

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2005, 15:47:38 ; Search time 28553 Seconds
(without alignments)
11865.614 Million cell updates/sec

Title: US-09-767-041-9

Perfect score: 6992

Sequence: 1 atcgccaaacgaattggca.....gcaaacagcaaaagaagctt 6992

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_btg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|--------------------|
| 1 | 6992 | 100.0 | 6992 | 1 | AF155804 Streptoco |
| 2 | 1195.8 | 17.1 | 15401 | 1 | AF118389 Streptoco |
| 3 | 955.6 | 13.7 | 17468 | 1 | AY057915 Streptoco |
| 4 | 955.6 | 13.7 | 17468 | 6 | AX283715 Sequence |
| 5 | 892.6 | 12.8 | 21562 | 1 | AX289861 Streptoco |
| 6 | 885.6 | 12.7 | 14943 | 1 | AX5787 S.pneumonia |
| 7 | 839.6 | 12.0 | 14202 | 1 | AX376403 Streptoco |
| 8 | 798.8 | 11.4 | 18239 | 1 | AF349539 Streptoco |
| 9 | 798.8 | 11.4 | 21365 | 1 | AE014245 Streptoco |
| 10 | 798.8 | 11.4 | 34980 | 6 | CQ655070 Sequence |
| 11 | 798.8 | 11.4 | 34980 | 6 | AX954530 Sequence |
| 12 | 743.4 | 10.6 | 17596 | 1 | AF355776 Streptoco |
| 13 | 729.4 | 10.4 | 4513 | 1 | AF332894 Streptoco |
| 14 | 729.4 | 10.4 | 4513 | 1 | AF363060 Streptoco |
| 15 | 729.4 | 10.4 | 4522 | 1 | AF332914 Streptoco |
| 16 | 729.4 | 10.4 | 4888 | 1 | AF332893 Streptoco |
| 17 | 729.4 | 10.4 | 4888 | 1 | AF363058 Streptoco |
| 18 | 727.8 | 10.4 | 4888 | 1 | AF363057 Streptoco |
| 19 | 727.8 | 10.4 | 4888 | 1 | AF363059 Streptoco |

| | | | | | | |
|----|-------|------|--------|---|-----------|-----------|
| 20 | 727.8 | 10.4 | 6865 | 6 | BD166249 | Beta-1,3- |
| 21 | 727.8 | 10.4 | 9987 | 1 | AB050723 | Streptoco |
| 22 | 726.8 | 10.4 | 4768 | 1 | AF332896 | Streptoco |
| 23 | 726.8 | 10.4 | 4768 | 1 | AF363055 | Streptoco |
| 24 | 726.8 | 10.4 | 4768 | 1 | AF363056 | Streptoco |
| 25 | 726.8 | 10.4 | 17276 | 1 | AF163833 | Streptoco |
| 26 | 725.2 | 10.4 | 4411 | 1 | AF332897 | Streptoco |
| 27 | 725.2 | 10.4 | 4411 | 1 | AF332898 | Streptoco |
| 28 | 725.2 | 10.4 | 4411 | 1 | AF332899 | Streptoco |
| 29 | 725.2 | 10.4 | 4411 | 1 | AF332900 | Streptoco |
| 30 | 725.2 | 10.4 | 95596 | 6 | AX602206 | Sequence |
| 31 | 725.2 | 10.4 | 128050 | 1 | SAG766849 | Streptoco |
| 32 | 724 | 10.4 | 4521 | 1 | AF332895 | Streptoco |
| 33 | 723.6 | 10.3 | 13421 | 1 | AY375362 | Streptoco |
| 34 | 721.4 | 10.3 | 10206 | 1 | AB017355 | Streptoco |
| 35 | 721.4 | 10.3 | 25021 | 1 | AB028896 | Streptoco |
| 36 | 720.4 | 10.3 | 2815 | 1 | AF332901 | Streptoco |
| 37 | 720.4 | 10.3 | 2815 | 1 | AF332902 | Streptoco |
| 38 | 714 | 10.2 | 16448 | 1 | AF337958 | Streptoco |
| 39 | 652.4 | 9.3 | 2217 | 1 | AF367973 | Streptoco |
| 40 | 652.4 | 9.3 | 2220 | 1 | AF363035 | Streptoco |
| 41 | 652.4 | 9.3 | 2226 | 1 | AF363032 | Streptoco |
| 42 | 652.4 | 9.3 | 2226 | 1 | AF363033 | Streptoco |
| 43 | 652.4 | 9.3 | 2226 | 1 | AF363034 | Streptoco |
| 44 | 650.8 | 9.3 | 2217 | 1 | AF363038 | Streptoco |
| 45 | 650.8 | 9.3 | 2217 | 1 | AF363039 | Streptoco |

ALIGNMENTS

| | | | | | | | | |
|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-----|--------|-----------------|--|--|--|
| RESULT 1 | AF155804 | 6992 bp | DNA | linear | BCT 19-DEC-1999 | | | |
| LOCUS | AF155804 | 6992 bp | DNA | linear | BCT 19-DEC-1999 | | | |
| DEFINITION | Streptococcus suis strain 6555 CpsIE (cpsIE) gene, partial cds; Cps2F (cps2F), CpsIG (cpsIG), CpsIH (cpsIH), CpsII (cpsII), and CpsIJ (cpsIJ) genes, complete cds; and CpsIK (cpsIK) gene, partial cds. | | | | | | | |
| ACCESSION | AF155804 | | | | | | | |
| VERSION | AF155804.1 | GI:6601338 | | | | | | |
| KEYWORDS | Streptococcus suis | | | | | | | |
| SOURCE | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; | | | | | | | |
| ORGANISM | Streptococcus. | | | | | | | |
| REFERENCE | 1 (bases 1 to 6992) | | | | | | | |
| AUTHORS | Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F., Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A. | | | | | | | |
| TITLE | Identification and characterization of the cps locus of Streptococcus suis serotype 2: the capsule protects against phagocytosis and is an important virulence factor | | | | | | | |
| JOURNAL | Infect. Immun. 67 (4), 1750-1756 (1999) | | | | | | | |
| MEDLINE | 99184998 | | | | | | | |
| PUBMED | 10085014 | | | | | | | |
| REFERENCE | 2 (bases 1 to 6992) | | | | | | | |
| AUTHORS | Smith,H.E., Veenbergen,V., van der Velde,J., Damman,M., Wisselink,H.J. and Smits,M.A. | | | | | | | |
| TITLE | The cps genes of Streptococcus suis serotypes 1, 2, and 9: development of rapid serotype-specific PCR assays | | | | | | | |
| JOURNAL | J. Clin. Microbiol. 37 (10), 3146-3152 (1999) | | | | | | | |
| MEDLINE | 99419130 | | | | | | | |
| PUBMED | 10488168 | | | | | | | |
| REFERENCE | 3 (bases 1 to 6992) | | | | | | | |
| AUTHORS | Smith,H.E., Veenbergen,V., Van der Velde,J., Damman,M., Wisselink,H.J. and Smits,M.A. | | | | | | | |
| TITLE | Direct Submission | | | | | | | |
| JOURNAL | Submitted (02-JUN-1999) Bacteriology, Animal Science and Health, P.O. Box 65, Lelystad 8200 AB, The Netherlands | | | | | | | |
| FEATURES | Location/Qualifiers | | | | | | | |
| source | 1. .6992 | | | | | | | |
| | /organism="Streptococcus suis" | | | | | | | |
| | /mol_type="genomic DNA" | | | | | | | |
| | /strain="6555" | | | | | | | |


```
gene
/db_xref="taxon:1307"
<1..1367
/genes="cps1E"
CDS
<1..1367
/genes="cps1E"
/notes="putative glycosyltransferase"
/codon_start=3
/transl_table=11
/product="Cps1E"
/protein_id="AAFI8940.1"
/db_xref="GI:6601343"
/translation="ROTKLAFOMIAVAISALITSHIPNADLNRSGIFIMVHYFAP
FISMPVEFYRGNLIEFEKTFNYSIIPLFTAVSFLENFALSRGAGVFTLLNF
VLVEFNVIKQKPSLFTIYOKKTLITTAERWENMQVLFESHKIQKNLVALV
LGEIDKINLSLPLYSVERAIBFSTREVVDHFINLPSEFLDKQFVSDFELLGIDV
SVINDSPGFTLAKKIKQLLGDHSIVTFSTNFYKPSHIMMKRLDILGAVVGLICGI
VSIILVPIIRDDGPAIFAOKRYGONGRIPTFYKFSMVYDAERKOLLSONOMQGM
KRFMKGTILLEQLDISYAKTSLDELPOFYNVILGMSLVGTRPTVDFEYKTFQG
KRLSFKPGITGLWQVSGRSNITDFDDVVRDLDAIYIDNWTIWSIDIKILLKTVKVVLLR
EGSK"
1374..1823
/genes="cps2F"
CDS
1374..1823
/genes="cps2F"
/codon_start=1
/transl_table=11
/product="Cps2F"
/protein_id="AAFI8941.1"
/db_xref="GI:6601340"
/translation="MKVCLVSGSGHLTHLYLLKLPWKBEERFWVTFKEDARSLKN
EKMPYCPPTNRLNINLVKTFPLAKILRDEKPDVLISSGAVAVPPFYIGKUFGAKT
IYIEVDFRVNKSILTGLKLVYPVTDIFIVQWEEKVYKPSINLGSIF"
1823..2317
/genes="cps1G"
CDS
1823..2317
/genes="cps1G"
/notes="putative glycosyltransferase"
/codon_start=1
/transl_table=11
/product="Cps1G"
/protein_id="AAFI8942.1"
/db_xref="GI:6601341"
/translation="MIPVTGTHEQQPNRLIKEIDLKNGSITDEIFIQTGYSDVIP
EYCKYKFLSYKMEQYVINKSEVICHGGPATFMSLSKGKQLLPFRQKQYGEHVD
HOVEFVRRLIQDNNILFIENIDDLFEKILIEVSKQTNFTSNNPFCEQLKQIVKFNED
QENE"
3036..4202
/genes="cps1H"
CDS
3036..4202
/genes="cps1H"
/notes="putative capsular polysaccharide polymerase"
/codon_start=1
/transl_table=11
/product="Cps1H"
/protein_id="AAFI8943.1"
/db_xref="GI:6601342"
/translation="MFKLFYDPPEYFIFKYFWLIIFPEQYVFLIFMNLILFHIFK
LTKYLILKNLILLFLMSILCFVSVTSMVEINFERLFADEFTAPIIWIITAIWYNLY
SFNIDYKLNKISFTSPFLVLLGLISALYIIQNGKDIVFLDRHLIGLDYLTGVKTRLY
GWNYPNTNTTIVSILPIFALIKNQMQFPFLCLAFIPYILSGSRISLSLAILII
CLWRYIGGFANIKKLIVIFVILLIILNTELUYHEILAVINRESNEARFIILQGS
IDKVLNNILFGYGISYSVTGWLGSHSYISFFYKSGIVGLILMLFSFYIRKSY
GVNGETALFFYTSIAIFPIYETIDPIIIILVLVLFSSIGIWNINFNKDMETKNE"
4195..5163
/genes="cps1I"
CDS
4195..5163
/genes="cps1I"
/notes="putative glycosyltransferase"
/codon_start=1
/transl_table=11
/product="Cps1I"
/protein_id="AAFI8944.1"
```

```
gene
/db_xref="GI:6601343"
/translation="MNDLISVIVPIINVQDYLKDCINSIINQYTNLEVLVNDGSTD
DSEKICLNTYMKNDGRIKYYKINGGLADARNFGLHATGKYIAFVSDDDYIEVAMFER
MDNITEYNADIAEIDFCLVDENGYTKKRNNSFHLVLTRETYKEFLSGSNIENNVMC
KLYSRDIKIDIKFOINNRSGIGEDLLFNLVNVTWVDTRVYNYVIRNSSLINQ
KFSINNIDLVRLNYPFKLGRFESHVFAKVIKEKVKCLNKWYSTDCLDNELPILE
SYRKEIRRPFFIKAKRYLSRKHLVTLVLMKFSPLVMLYKFKQK"
5172..6143
/genes="cps1J"
CDS
5172..6143
/genes="cps1J"
/notes="putative glycosyltransferase"
/codon_start=1
/transl_table=11
/product="Cps1J"
/protein_id="AAFI8945.1"
/db_xref="GI:6601344"
/translation="MDKISVIVPVNVVVKYLSKCSIESIINQYKNIEILLIDGSDVD
SAKICKYAEKQKRVKIFFTNHSGVSNARNHGKIRTAETVIMFVSDDDVDSRLVEKL
YFNITKSRDLSGLYATFSENINNFNNIDFEAINTVQDMKEKFNMLYINNIF
STPVCKYKRYITDLFQENOMLGEDLLFNHLKNDIRVSYLTHELYFYRRGILSTV
NSFKGVFLQLENLQKQVILFKQIYGEDPDVSIKVDITRMQVFFYSLLMFYKQKSI
FDKFLIFRNLKYKYFNLLKVSNSLSKNFCIRIVSNKVKILML"
6156..6992
/genes="cps1K"
CDS
6156..6992
/genes="cps1K"
/notes="putative glycosyltransferase"
/codon_start=1
/transl_table=11
/product="Cps1K"
/protein_id="AAFI8946.1"
/db_xref="GI:6601345"
/translation="MDTISKISIVPIYVNVKYLKSCIDSIVNQTYKHIEILLVNDGS
TDNBERICLAYAKDSIRIRYPKKGGLSDARNYGSIRAKGDYLAIFDSDDFTHSFI
ORLHEAIERENALVAGVDRVDASGHFTABPLTNQAVLSCGVCKLLEADGHRF
VACNKLYKELFEDFRPEKGIHEDEYFTYRLLYLEKVAIVKELCYYYVDRENSIT
TSSMTDHRPHCLLEFQNERMDFYESRDKLELLECYRSLAFVFLGKYNHWSLQKQ
KKL"
```

ORIGIN

| Query Match | 100.0% | Score 6992 | DB 1 | Length 6992 |
|-----------------------|----------------|--------------------------------------------------------------|----------|-------------|
| Best Local Similarity | 100.0% | Pred. No. 0 | | |
| Matches 6992 | Conservative 0 | Mismatches 0 | Indels 0 | Gaps 0 |
| Qy | 1 | ATCGCAAAACGAAATGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAA | 60 | |
| Db | 1 | ATCGCAAAACGAAATGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAA | 60 | |
| Qy | 61 | CAAGTCATATACCAATGCTGATTAATAATCGTTCTGGAATTTTATCATATGATGTTTC | 120 | |
| Db | 61 | CAAGTCATATACCAATGCTGATTAATAATCGTTCTGGAATTTTATCATATGATGTTTC | 120 | |
| Qy | 121 | ATTATTTTGCATTTTATATCTCGTATGCCAGTTGCAATTTGAGTATAGAGGTAATCTGA | 180 | |
| Db | 121 | ATTATTTTGCATTTTATATCTCGTATGCCAGTTGCAATTTGAGTATAGAGGTAATCTGA | 180 | |
| Qy | 181 | TAGAGTTTGAAAAACATTTAACTAGTATATATTTGCAATTTTCTTACGGCAGTAT | 240 | |
| Db | 181 | TAGAGTTTGAAAAACATTTAACTAGTATATATTTGCAATTTTCTTACGGCAGTAT | 240 | |
| Qy | 241 | CATTTTGTGGAGATAATTTTCGCACTTTCAAGACGTGGTCCGTGATTTCACATTA | 300 | |
| Db | 241 | CATTTTGTGGAGATAATTTTCGCACTTTCAAGACGTGGTCCGTGATTTCACATTA | 300 | |
| Qy | 301 | TAAACTTCGTTTGGTATACCTATTAACTAATTAATTAAGCAGTTTAAGGATGCTTTC | 360 | |
| Db | 301 | TAAACTTCGTTTGGTATACCTATTAACTAATTAATTAAGCAGTTTAAGGATGCTTTC | 360 | |
| Qy | 361 | TATTTTCGACAACTCTATCAAAAAAGACGATTTCTAATTAACACGGCTGACGATGGGAAA | 420 | |
| Db | 361 | TATTTTCGACAACTCTATCAAAAAAGACGATTTCTAATTAACACGGCTGACGATGGGAAA | 420 | |

Qy 421 ATATGCAAGTTTTATTTGAATCACAATAAACAATTCAAAAATCTTGTGCAATTTGGTAG 480
Db 421 ATATGCAAGTTTTATTTGAATCACAATAAACAATTCAAAAATCTTGTGCAATTTGGTAG 480
Qy 481 TTTTGGGTACAGAAATAGATAAAATTAATTTATCAATTACCGCTCTATTATTCTGTGGAAG 540
Db 481 TTTTGGGTACAGAAATAGATAAAATTAATTTATCAATTACCGCTCTATTATTCTGTGGAAG 540
Qy 541 AAGCTATAGAGTTTTCACAAGGGAAGTGGTCGACCACGCTCTTTATAAATCTACCAAGTG 600
Db 541 AAGCTATAGAGTTTTCACAAGGGAAGTGGTCGACCACGCTCTTTATAAATCTACCAAGTG 600
Qy 601 AGTTTTAGAGCTAAAGCAATTCGTTTTAGAGTTTGGAGTTTGGTAGTTTGGTAGCGG 660
Db 601 AGTTTTAGAGCTAAAGCAATTCGTTTTAGAGTTTGGAGTTTGGTAGTTTGGTAGCGG 660
Qy 661 TTGATATTAATTCATTCGGTTTTTCTCGTTTGAAACCAAAAAATCAAACTGCTAGGTG 720
Db 661 TTGATATTAATTCATTCGGTTTTTCTCGTTTGAAACCAAAAAATCAAACTGCTAGGTG 720
Qy 721 ACCATAGCATTTGAATCTTTTCCACAAATTTTATAAGCCTAGTCATATCATGATGAAC 780
Db 721 ACCATAGCATTTGAATCTTTTCCACAAATTTTATAAGCCTAGTCATATCATGATGAAC 780
Qy 781 GACTTTTGGATATATCTCGAGCGGTAGTCGGGTAAATTTATTTGGTATAGTTTCTATT 840
Db 781 GACTTTTGGATATATCTCGAGCGGTAGTCGGGTAAATTTATTTGGTATAGTTTCTATT 840
Qy 841 TGTGTAGTCCAAATTTATTCGTAGAGTGGTGACCGGCTATTCTTCTCAGAAAACGAGTTG 900
Db 841 TGTGTAGTCCAAATTTATTCGTAGAGTGGTGACCGGCTATTCTTCTCAGAAAACGAGTTG 900
Qy 901 GACAGATGGACGATATTTACATTTCAAGTTTCGATCGATGATGTCGTCGAGG 960
Db 901 GACAGATGGACGATATTTACATTTCAAGTTTCGATCGATGATGTCGTCGAGG 960
Qy 961 AGCCCAAAAAAGACTGTGCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTTAAAAATGG 1020
Db 961 AGCCCAAAAAAGACTGTGCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTTAAAAATGG 1020
Qy 1021 GAAAAACGATCCTAGAAATTTACTCCAAATTTGGACATTTTCATAGCAGAAAACAGTTTAGACG 1080
Db 1021 GAAAAACGATCCTAGAAATTTACTCCAAATTTGGACATTTTCATAGCAGAAAACAGTTTAGACG 1080
Qy 1081 AGTTTACCAAGTTTTATTAATGTTTTAAATTTGCGATATGAGTCTAGTTGCTACAGTCCAC 1140
Db 1081 AGTTTACCAAGTTTTATTAATGTTTTAAATTTGCGATATGAGTCTAGTTGCTACAGTCCAC 1140
Qy 1141 CTACAGTTTGATGAATTTTGAAAAATATATCTCCTGGTCAAAAGAGACGATTTGAGTTTAAAC 1200
Db 1141 CTACAGTTTGATGAATTTTGAAAAATATATCTCCTGGTCAAAAGAGACGATTTGAGTTTAAAC 1200
Qy 1201 CAGGGAATTAAGGTCTCTGGCAGGTAGTGGTCTGATGATTAATATCAAGCTTCGACGACG 1260
Db 1201 CAGGGAATTAAGGTCTCTGGCAGGTAGTGGTCTGATGATTAATATCAAGCTTCGACGACG 1260
Qy 1261 TAGTTCCGTTGCACTTAGCATATGATTAATTCGACTATCTGGTCAAGATTTAAAAATTT 1320
Db 1261 TAGTTCCGTTGCACTTAGCATATGATTAATTCGACTATCTGGTCAAGATTTAAAAATTT 1320
Qy 1321 TATTAAAGACAGTGAAGTTGATTTGAGAGAGGGAAGTAAAGTAAAGTATATGAAG 1380
Db 1321 TATTAAAGACAGTGAAGTTGATTTGAGAGAGGGAAGTAAAGTAAAGTATATGAAG 1380
Qy 1381 TTTGTTTGGTCCGTTCTTCAGGGGACATTTGACTCACTGTTGATTTGTTAAAAACCGTTTT 1440
Db 1381 TTTGTTTGGTCCGTTCTTCAGGGGACATTTGACTCACTGTTGATTTGTTAAAAACCGTTTT 1440
Qy 1441 GGAAGGAAGAAGAGCTTTTGGGTAAACATTTGATAAAGAGGATGCAAGAGTCTTTTGA 1500
Db 1441 GGAAGGAAGAAGAGCTTTTGGGTAAACATTTGATAAAGAGGATGCAAGAGTCTTTTGA 1500
Qy 1501 AGAATGAAAAAATGATCCATGTTACTTTTCCAAACAATCGCAATCTCATTAATTTAGTGA 1560

Db 1501 AGAATGAAAAAATGATCCATGTTACTTTCCAAACAATCGCAATCTCATTAATTTAGTGA 1560
Qy 1561 AAAATACTTTCTTAGCTTTTCAAAAATTTTACGTGATGAGAAACCAAGATGTTATTTTCAT 1620
Db 1561 AAAATACTTTCTTAGCTTTTCAAAAATTTTACGTGATGAGAAACCAAGATGTTATTTTCAT 1620
Qy 1621 CTGTGCGCGCGTTGCTGTCCTCTTTTATCATCGGAAAACTATTTGGAGCAAGACGA 1680
Db 1621 CTGTGCGCGCGTTGCTGTCCTCTTTTATCATCGGAAAACTATTTGGAGCAAGACGA 1680
Qy 1681 TTTTATTTGAAGTATTTGATCGAGTTAATAAATCTACATTAACCTGGAATACTAGTTTATC 1740
Db 1681 TTTTATTTGAAGTATTTGATCGAGTTAATAAATCTACATTAACCTGGAATACTAGTTTATC 1740
Qy 1741 CCGTAAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGTATATCTTAATCTA 1800
Db 1741 CCGTAAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGTATATCTTAATCTA 1800
Qy 1801 TTAACTTTGGGAGTATTTTAAATGATTTTGTAACTAGTAGGAACTCATGAACACAGTT 1860
Db 1801 TTAACTTTGGGAGTATTTTAAATGATTTTGTAACTAGTAGGAACTCATGAACACAGTT 1860
Qy 1861 TAATCGATTGNTAAAGAGATTTGATTTTGAAGAAATGGAAGTATAACCGAGGAAT 1920
Db 1861 TAATCGATTGNTAAAGAGATTTGATTTTGAAGAAATGGAAGTATAACCGAGGAAT 1920
Qy 1921 ATTTTATTTCAACACGATATTTCTGACTATATTTCCAGAAATATTGCAAGTATAAAAAATTTCT 1980
Db 1921 ATTTTATTTCAACACGATATTTCTGACTATATTTCCAGAAATATTGCAAGTATAAAAAATTTCT 1980
Qy 1981 CAGTTACAAAGAAATGGAACAAATATATTAACAATCAGAAGTAGTTATTTGCCACGAGG 2040
Db 1981 CAGTTACAAAGAAATGGAACAAATATATTAACAATCAGAAGTAGTTATTTGCCACGAGG 2040
Qy 2041 CCCCCTACTTTTATGAATTCATTATCCAAAGGAAAAAACAATTTATTTGTTCTCTAGACA 2100
Db 2041 CCCCCTACTTTTATGAATTCATTATCCAAAGGAAAAAACAATTTATTTGTTCTCTAGACA 2100
Qy 2101 AAAAAAGTATGGTGAACAATGTAATGATCATCAAGTAGAGTTTGTGAAGAAATTTTACA 2160
Db 2101 AAAAAAGTATGGTGAACAATGTAATGATCATCAAGTAGAGTTTGTGAAGAAATTTTACA 2160
Qy 2161 AGATAATAATATTTTATTTATAGAAAAATATAGATGATTTGTTGAAAAATTTATTTGAAGT 2220
Db 2161 AGATAATAATATTTTATTTATAGAAAAATATAGATGATTTGTTGAAAAATTTATTTGAAGT 2220
Qy 2221 TTCTAAGCAAACTAACTTTTACATCAAAATTAATTTTGTGTGAAGATTTAAAAACAAT 2280
Db 2221 TTCTAAGCAAACTAACTTTTACATCAAAATTAATTTTGTGTGAAGATTTAAAAACAAT 2280
Qy 2281 AGTTGAAAAATTTTAACTGAGGATCAAGAAATGAATTAATAAAGATGATATTTGATAA 2340
Db 2281 AGTTGAAAAATTTTAACTGAGGATCAAGAAATGAATTAATAAAGATGATATTTGATAA 2340
Qy 2341 TGGCTTATCAATAATTTTCTCAGATTTTACCTGGAGAGGATACAGATATTTATCTTCT 2400
Db 2341 TGGCTTATCAATAATTTTCTCAGATTTTACCTGGAGAGGATACAGATATTTATCTTCT 2400
Qy 2401 CTCAGGGAATGCAACCATTTAGTTTCCCTTCAGAAATACCTGTATATTTATTTAAATATTC 2460
Db 2401 CTCAGGGAATGCAACCATTTAGTTTCCCTTCAGAAATACCTGTATATTTATTTAAATATTC 2460
Qy 2461 TCAGGATTTATATGTTGATTTTACAAAGATGAGCAAAATATAAAGAAATAGATATA 2520
Db 2461 TCAGGATTTATATGTTGATTTTACAAAGATGAGCAAAATATAAAGAAATAGATATA 2520
Qy 2521 TGAACGAGTTTAAATGTTTACAGATTTATTTCCCTAAATATATCAGAAAAAATCTATTGATAATGT 2580
Db 2521 TGAACGAGTTTAAATGTTTACAGATTTATTTCCCTAAATATATCAGAAAAAATCTATTGATAATGT 2580
Qy 2581 ACTGTTTAGAAATTTTATTTAAGAATGATTCGAGCTTTTCAATCTATTTTCAAGAGTTGTT 2640
Db 2581 ACTGTTTAGAAATTTTATTTAAGAATGATTCGAGCTTTTCAATCTATTTTCAAGAGTTGTT 2640

Db 2581 ACTGTTTAGAATTTTATTAAAGATGTATCGAGCTTTTGAATPACTATTTACAAAGATTGTT 2640

Qy 2641 GTTTATTGATAGATAAAAAACATGGCTTAAGAATAAGATTGGTCTTAATTTGGGTTTCG 2700

Db 2641 GTTTATTGATAGATAAAAAACATGGCTTAAGAATAAGATTGGTCTTAATTTGGGTTTCG 2700

Qy 2701 CTTCCACATGATTTTGTGGCAATTCCTTTTATCAAAATGAAGCAAGAAACAGCTTTATTTATTT 2760

Db 2701 CTTCCACATGATTTTGTGGCAATTCCTTTTATCAAAATGAAGCAAGAAACAGCTTTATTTATTT 2760

Qy 2761 AAGTAAATCTAAATCTGTCAGATGAATCTATTTATATACAGACAAATTTAGAAAATATGAATTT 2820

Db 2761 AAGTAAATCTAAATCTGTCAGATGAATCTATTTATATACAGACAAATTTAGAAAATATGAATTT 2820

Qy 2821 TCAAAATAGATTTATCTAAATATGGAATTTTAAAGATATATAAAGTGGAAAAAATCAACATCT 2880

Db 2821 TCAAAATAGATTTATCTAAATATGGAATTTTAAAGATATATAAAGTGGAAAAAATCAACATCT 2880

Qy 2881 TCTCCTATTGCTTTTACAGATGATCTTATGATGAATTTGCTTAAGATATATAAAGTGGAAAAATTTAGGT 2940

Db 2881 TCTCCTATTGCTTTTACAGATGATCTTATGATGAATTTGCTTAAGATATATAAAGTGGAAAAATTTAGGT 2940

Qy 2941 TTTTATTGCTAGAAAGTTTAAAAATAGAAAATAAAATCTAAATTTAAAGAAAATTTATTAAT 3000

Db 2941 TTTTATTGCTAGAAAGTTTAAAAATAGAAAATAAAATCTAAATTTAAAGAAAATTTATTAAT 3000

Qy 3001 AAAAAATPAAATAGTTGATTTTGTGAGAGTAATGTATGTTTAAATTTTAAATATTAATATGACC 3060

Db 3001 AAAAAATPAAATAGTTGATTTTGTGAGAGTAATGTATGTTTAAATTTTAAATATTAATATGACC 3060

Qy 3061 CGGAATATTTATTTTAAAGTACTCTGTTGATTTTATTTTCCAGAGCAAAAGTATG 3120

Db 3061 CGGAATATTTATTTTAAAGTACTCTGTTGATTTTATTTTCCAGAGCAAAAGTATG 3120

Qy 3121 TATTTTATTAAATTTTATGAAATTTAAATTTTATTTTATATATAAAATTTTGAAGAACTAAGC 3180

Db 3121 TATTTTATTAAATTTTATGAAATTTAAATTTTATTTTATATATAAAATTTTGAAGAACTAAGC 3180

Qy 3181 TAATATTAAAAATGAAATTTTATGTTTATTTATATGCTATATATGTTTGTGTTTTCAG 3240

Db 3181 TAATATTAAAAATGAAATTTTATGTTTATTTATATGCTCTATATATGTTTGTGTTTTCAG 3240

Qy 3241 TAGTCACAAGTATGTTTGTGAAATAAATTTTGAAGATTTATTTGCGATTTTACTGCTC 3300

Db 3241 TAGTCACAAGTATGTTTGTGAAATAAATTTTGAAGATTTATTTGCGATTTTACTGCTC 3300

Qy 3301 CCATAATTTGGAATTTATGCAATATGTAATTTATATATTTGTAATTTTAAATATTTGATT 3360

Db 3301 CCATAATTTGGAATTTATGCAATATGTAATTTATATATTTGTAATTTTAAATATTTGATT 3360

Qy 3361 ATAAAAATTTAAAAATPAGTATCTTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTGAT 3420

Db 3361 ATAAAAATTTAAAAATPAGTATCTTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTGAT 3420

Qy 3421 TGATATTTATTTCAAAATGGAAGATATGTAATTTTATGAGTATGTAATTTTATGAGTATGTAATTT 3480

Db 3421 TGATATTTATTTCAAAATGGAAGATATGTAATTTTATGAGTATGTAATTTTATGAGTATGTAATTT 3480

Qy 3481 ACTATCTTATAACAGGCGTCAAAACAGAGTTGGTTGGCTTTATGAACTATCTCACTACGTTAA 3540

Db 3481 ACTATCTTATAACAGGCGTCAAAACAGAGTTGGTTGGCTTTATGAACTATCTCACTACGTTAA 3540

Qy 3541 ATACCACACAAATPATAATAGTTTCAATTCGGTTTAAATCTTTGCACTTATAAAAAATAAATGC 3600

Db 3541 ATACCACACAAATPATAATAGTTTCAATTCGGTTTAAATCTTTGCACTTATAAAAAATAAATGC 3600

Qy 3601 AACAAATTTTCTTCTGCTGCTTTTATACCGATCTATTTAAGTGGATCCGAGATTG 3660

Db 3601 AACAAATTTTCTTCTGCTGCTTTTATACCGATCTATTTAAGTGGATCCGAGATTG 3660

Qy 3661 GTAGTTTATCGTAGCAATATTAATATATGCTTTGTTATGAGATATATAGGTGGAAAAAT 3720

Db 3661 GTAGTTTATCGTAGCAATATTAATATATGCTTTGTTATGAGATATATAGGTGGAAAAAT 3720

Qy 3721 TTGCTTGGATPAAAAAGCTAATAGTAAATATTTGTAATPACTATTTATTTAAATACTG 3780

Db 3721 TTGCTTGGATPAAAAAGCTAATAGTAAATATTTGTAATPACTATTTATTTAAATACTG 3780

Qy 3781 AATTGCTTTTACCAATGAAATTTTGGCTGTTTATTAATTTCTAGAGATCAAGTAACCAAGCTA 3840

Db 3781 AATTGCTTTTACCAATGAAATTTTGGCTGTTTATTAATTTCTAGAGATCAAGTAACCAAGCTA 3840

Qy 3841 GATTTATTTATTTATCAAGGAAGTATTGATAAAGTATTAGAAAAACAATTTATTTTGGAT 3900

Db 3841 GATTTATTTATTTATCAAGGAAGTATTGATAAAGTATTAGAAAAACAATTTATTTTGGAT 3900

Qy 3901 ATGGAATATCCGAATATTTCAAGTTACGGGAACCTTGGCTCGGAAGTCAATTCAGGCTATATAT 3960

Db 3901 ATGGAATATCCGAATATTTCAAGTTACGGGAACCTTGGCTCGGAAGTCAATTCAGGCTATATAT 3960

Qy 3961 CATTTTATTAATATCAGGAATAGTTGGGTTGATTTTACTGATGTTTCTTTTATG 4020

Db 3961 CATTTTATTAATATCAGGAATAGTTGGGTTGATTTTACTGATGTTTCTTTTATG 4020

Qy 4021 TTATAAAAAAGCTTATGGAAGTTAAATGGGGAAAAACAGCACTATTTTATTTACATCATTAG 4080

Db 4021 TTATAAAAAAGCTTATGGAAGTTAAATGGGGAAAAACAGCACTATTTTATTTACATCATTAG 4080

Qy 4081 CCATATTTTTCATATATGAAAACAATAGATCCGATTTATTTATATATTTAGTACTATCTTTT 4140

Db 4081 CCATATTTTTCATATATGAAAACAATAGATCCGATTTATTTATATATTTAGTACTATCTTTT 4140

Qy 4141 CTTCAATAGTATTTGGAATAATATAAATTTTAAAAAGGATATGGAGACAAAAAATCAAT 4200

Db 4141 CTTCAATAGTATTTGGAATAATATAAATTTTAAAAAGGATATGGAGACAAAAAATCAAT 4200

Qy 4201 GATTTAAATTTTCAAGTATTTGACCAATTTTATAATGTCACCAATTTATCTTGATAATGTTATT 4260

Db 4201 GATTTAAATTTTCAAGTATTTGACCAATTTTATAATGTCACCAATTTATCTTGATAATGTTATT 4260

Qy 4261 AACAGTATTTATTAACCAACATATATCTAAATTTAGAGGTTATTTCTCGTAAATGATGGAAGT 4320

Db 4261 AACAGTATTTATTAACCAACATATATCTAAATTTAGAGGTTATTTCTCGTAAATGATGGAAGT 4320

Qy 4321 ACTGATGATTTCTGAGAAAAATTTGCTTTAACTATATGAAGAACGATGGAAGAAATTTAAATAT 4380

Db 4321 ACTGATGATTTCTGAGAAAAATTTGCTTTAACTATATGAAGAACGATGGAAGAAATTTAAATAT 4380

Qy 4381 TACAAGAAAAATTAATCGCGGCTTAGCAGATGCTCGAAATTTTCGACCTAGAACATGCAACA 4440

Db 4381 TACAAGAAAAATTAATCGCGGCTTAGCAGATGCTCGAAATTTTCGACCTAGAACATGCAACA 4440

Qy 4441 GGTAAATATATTTGCTTTGTCGATTTCTGATGACTATATAGAAAGTTGCAATGTTTCGAGAGA 4500

Db 4441 GGTAAATATATTTGCTTTGTCGATTTCTGATGACTATATAGAAAGTTGCAATGTTTCGAGAGA 4500

Qy 4501 ATCSATGATAATAATACTGAGTATAATCCGATATACGAGATAGATTTTGTGTTAGTA 4560

Db 4501 ATCSATGATAATAATACTGAGTATAATCCGATATACGAGATAGATTTTGTGTTAGTA 4560

Qy 4561 GACGAAAAACGGGTATACAAAGAAAAAAGAAATAGTAAATTTTCATGTCCTTAAACGAGAGAA 4620

Db 4561 GACGAAAAACGGGTATACAAAGAAAAAAGAAATAGTAAATTTTCATGTCCTTAAACGAGAGAA 4620

Qy 4621 GAGACTGTAAAAAGAAATTTTGTGTCAGGATCTAAATATAGAAAAATAATGTTGTCGCAAGCTT 4680

Db 4621 GAGACTGTAAAAAGAAATTTTGTGTCAGGATCTAAATATAGAAAAATAATGTTGTCGCAAGCTT 4680

Qy 4681 TATTTCAGAGATATTAATAAGATATAAATTTCCAAATTAATAATAGAAATTAATGTTGGTCAAGCTT 4740

Db 4681 TATTTCAGAGATATTAATAAGATATAAATTTCCAAATTAATAATAGAAATTAATGTTGGTCAAGCTT 4740

Qy 4741 GATTTGCTTTTAAATTTGGAGGTCTTTGAAACAATGTAACAGCTGTAGTGTAGTACTAGTA 4800

Db 4741 GATTTGCTTTTAAATTTGGAGGTCTTTGAAACAATGTAACAGCTGTAGTGTAGTACTAGTA 4800

| | | | |
|----|------|---------------------------------------------------------------------|------|
| Qy | 4801 | GAATATTATTAATAATTATGTCATTCGTAAACAGTTTCGCTTATTAATCAAGAAATTCCTATATA | 4860 |
| Db | 4801 | GAATATTATTAATAATTATGTCATTCGTAAACAGTTTCGCTTATTAATCAAGAAATTCCTATATA | 4860 |
| Qy | 4861 | AATTAATATTGATTTAGTCACAGAAATTCAGAAATTCACCCCTTTAAAGTTTAAAGAGAGATTTT | 4920 |
| Db | 4861 | AATTAATATTGATTTAGTCACAGAAATTCAGAAATTCACCCCTTTAAAGTTTAAAGAGAGAGATTTT | 4920 |
| Qy | 4921 | AGTCATTATTTTGTGATGCAAAAGTTTAAAGAGAGAGTTTAAATGTTTAAACCAAAATGTTAT | 4980 |
| Db | 4921 | AGTCATTATTTTGTGATGCAAAAGTTTAAAGAGAGAGTTTAAATGTTTAAACCAAAATGTTAT | 4980 |
| Qy | 4981 | TCAACAGATTTGTTGGATAATGAGTTCTGTGCAATATTAAGAGTCTTATCGAAAGAGAAATA | 5040 |
| Db | 4981 | TCAACAGATTTGTTGGATAATGAGTTCTGTGCAATATTAAGAGTCTTATCGAAAGAGAAATA | 5040 |
| Qy | 5041 | CGTAGATATTCATTTTAAAGCGAAAGATATTTATCAAGAAAGCATTTAGTTACGTTG | 5100 |
| Db | 5041 | CGTAGATATTCATTTTAAAGCGAAAGATATTTATCAAGAAAGCATTTAGTTACGTTG | 5100 |
| Qy | 5101 | TATTTGATGAATTTTCGCCCTAAACTATATGTAATGTTTATTAAGAAATTTCAAAAGCAG | 5160 |
| Db | 5101 | TATTTGATGAATTTTCGCCCTAAACTATATGTAATGTTTATTAAGAAATTTCAAAAGCAG | 5160 |
| Qy | 5161 | TAGAGGTAAGAAATCGATAAATTTAGTGTATTTGTTCCAGTTTATAAATGATAGATAATTT | 5220 |
| Db | 5161 | TAGAGGTAAGAAATCGATAAATTTAGTGTATTTGTTCCAGTTTATAAATGATAGATAATTT | 5220 |
| Qy | 5221 | TAGTAGTTGTATAGAAAGCATTTATTAATCAAAATTTATAAATATAGAAATATATTGA | 5280 |
| Db | 5221 | TAGTAGTTGTATAGAAAGCATTTATTAATCAAAATTTATAAATATAGAAATATATTGA | 5280 |
| Qy | 5281 | TAGATGATGGCTCTGTAGATGATCTGCTAAATATGCAAGAAATATCGCAAAAGAGATA | 5340 |
| Db | 5281 | TAGATGATGGCTCTGTAGATGATCTGCTAAATATGCAAGAAATATCGCAAAAGAGATA | 5340 |
| Qy | 5341 | AAAGAGTAAATTTTTTTTCACTAATCATAGTGGAGTATCAAAATGCTAGAAATCATGGAA | 5400 |
| Db | 5341 | AAAGAGTAAATTTTTTTTCACTAATCATAGTGGAGTATCAAAATGCTAGAAATCATGGAA | 5400 |
| Qy | 5401 | TAAAGCGAGTACAGCTGAATATATATGTTGTTGACTCTGATGATGATGTTGTTGATAGTA | 5460 |
| Db | 5401 | TAAAGCGAGTACAGCTGAATATATATGTTGTTGACTCTGATGATGATGTTGTTGATAGTA | 5460 |
| Qy | 5461 | GATTAGTAGAATAATTTATTTTAAATTTATTAATAATTAATAAAGTAGAGTATTTCTGCTGTT | 5520 |
| Db | 5461 | GATTAGTAGAATAATTTATTTTAAATTTATTAATAATTAATAAAGTAGAGTATTTCTGCTGTT | 5520 |
| Qy | 5521 | TGTACGCTACTTTTTCAGAAATATAAATAATTTTGAAGTGAATAATCCAAATATTGATT | 5580 |
| Db | 5521 | TGTACGCTACTTTTTCAGAAATATAAATAATTTTGAAGTGAATAATCCAAATATTGATT | 5580 |
| Qy | 5581 | TTGAAGCAATTAATACCGTGCAGGACATGGAGAGAAAAATTTTATGAATTTGTATATAA | 5640 |
| Db | 5581 | TTGAAGCAATTAATACCGTGCAGGACATGGAGAGAAAAATTTTATGAATTTGTATATAA | 5640 |
| Qy | 5641 | ATAATAATTTTCTACTCTGTTGTAACATATATAAGAAAGATACATAACAGATCTTT | 5700 |
| Db | 5641 | ATAATAATTTTCTACTCTGTTGTAACATATATAAGAAAGATACATAACAGATCTTT | 5700 |
| Qy | 5701 | TTCAAGAGAAATCAATGTTTGGAGAGATTTTACTTTTAAATCTGATTATTTAAAGAAATA | 5760 |
| Db | 5701 | TTCAAGAGAAATCAATGTTTGGAGAGATTTTACTTTTAAATCTGATTATTTAAAGAAATA | 5760 |
| Qy | 5761 | TAGATAGAGTTAGTTATTTGATCAATCTTTTATTTTATAGGAGAGGTATACCTAAGTA | 5820 |
| Db | 5761 | TAGATAGAGTTAGTTATTTGATCAATCTTTTATTTTATAGGAGAGGTATACCTAAGTA | 5820 |
| Qy | 5821 | CAGTAAATCTTTTAAAGAGGTGTTTTCGAAATTCGAAATTTGCAAAAACAGTGA | 5880 |
| Db | 5821 | CAGTAAATCTTTTAAAGAGGTGTTTTCGAAATTTGCAAAATTTGCAAAAACAGTGA | 5880 |
| Qy | 5881 | TAGTATTGTTTAAAGCAATATATGTTGAGGATTTTTCGATCATATTGTTAAAGATATA | 5940 |

| | | | |
|----|------|---------------------------------------------------------------|------|
| Db | 5881 | TAGTATTGTTTAAAGCAATATATGTTGAGGATTTTTCGATCATATTGTTTAAAGATATA | 5940 |
| Qy | 5941 | TACGTTGCGCAAGTATTTTATTTATAGCTTTTAAAGTTTAAATACGGAAGACAGTCTATTT | 6000 |
| Db | 5941 | TACGTTGCGCAAGTATTTTATTTATAGCTTTTAAAGTTTAAATACGGAAGACAGTCTATTT | 6000 |
| Qy | 6001 | TTGCAAAATTTTAAATTTTAAAGTCTTTTATATAAAAAATATTTTAACTTTGTTTAAAG | 6060 |
| Db | 6001 | TTGCAAAATTTTAAATTTTAAAGTCTTTTATATAAAAAATATTTTAACTTTGTTTAAAG | 6060 |
| Qy | 6061 | TATCTAACAAATTTCTTTGCTTAAATTTTCTGATAGAAATTTGTTTTCGAACAAAGTTT | 6120 |
| Db | 6061 | TATCTAACAAATTTCTTTGCTTAAATTTTCTGATAGAAATTTGTTTTCGAACAAAGTTT | 6120 |
| Qy | 6121 | TTAAAAAATATTTATGTTTAAATAGGAAGATATCATGGATACATTTAGTAAAAATTTCTA | 6180 |
| Db | 6121 | TTAAAAAATATTTATGTTTAAATAGGAAGATATCATGGATACATTTAGTAAAAATTTCTA | 6180 |
| Qy | 6181 | TAAATTTGACCTATATATAATAGTAAAAATTTTATCTAAATGATAGATAGCATTTGTA | 6240 |
| Db | 6181 | TAAATTTGACCTATATATAATAGTAAAAATTTTATCTAAATGATAGATAGCATTTGTA | 6240 |
| Qy | 6241 | ATCAGACCTACAAACATATAGAGATTTCTCTGTTGAATGACGGTAGTACGGATTAATTCGG | 6300 |
| Db | 6241 | ATCAGACCTACAAACATATAGAGATTTCTCTGTTGAATGACGGTAGTACGGATTAATTCGG | 6300 |
| Qy | 6301 | AAGAAAAATTTGTTAGCATATGCGAAGAAAGATAGTCGATTCGTTTATTTTAAAAAGAGA | 6360 |
| Db | 6301 | AAGAAAAATTTGTTAGCATATGCGAAGAAAGATAGTCGATTCGTTTATTTTAAAAAGAGA | 6360 |
| Qy | 6361 | ACGCGGGCTATCAGATGCGCGTAAATTTATGTCATAAGTTCGCGCAAGGGTGAATCTTAG | 6420 |
| Db | 6361 | ACGCGGGCTATCAGATGCGCGTAAATTTATGTCATAAGTTCGCGCAAGGGTGAATCTTAG | 6420 |
| Qy | 6421 | CTTTTATAGACTCAGATGATTTTATTCATTCCGAGTTTCATCCAGCTTTTACAGAGCA | 6480 |
| Db | 6421 | CTTTTATAGACTCAGATGATTTTATTCATTCCGAGTTTCATCCAGCTTTTACAGAGCA | 6480 |
| Qy | 6481 | TTGAGAGAGAAATGCGCTTGTGCGAGTTGCTGTTTATGATAGGTTAGATGCTTCGGGGC | 6540 |
| Db | 6481 | TTGAGAGAGAAATGCGCTTGTGCGAGTTGCTGTTTATGATAGGTTAGATGCTTCGGGGC | 6540 |
| Qy | 6541 | ATTTCTTAAACAGCAGAGCGCTTCTTACAAATCAGGCTGTTCTGAGCGGCAAGGAATGTTT | 6600 |
| Db | 6541 | ATTTCTTAAACAGCAGAGCGCTTCTTACAAATCAGGCTGTTCTGAGCGGCAAGGAATGTTT | 6600 |
| Qy | 6601 | GTAAGAAAGCTGCTAGAGCGGATGCTCATCGCTTTGTTGTTGCTGTTAAATCTCTATA | 6660 |
| Db | 6601 | GTAAGAAAGCTGCTAGAGCGGATGCTCATCGCTTTGTTGTTGCTGTTAAATCTCTATA | 6660 |
| Qy | 6661 | AAAAAGAACTATTTTGAAGATTTTTCGATTTGAAAAGGGTGAAGATTCATGAAGATGAAT | 6720 |
| Db | 6661 | AAAAAGAACTATTTTGAAGATTTTTCGATTTGAAAAGGGTGAAGATTCATGAAGATGAAT | 6720 |
| Qy | 6721 | TCATCTATCGCTCTCTATGAGTTAGAAAAGTTGCAATAGTTAAAGAGTGTGTTGACT | 6780 |
| Db | 6721 | TCATCTATCGCTCTCTATGAGTTAGAAAAGTTGCAATAGTTAAAGAGTGTGTTGACT | 6780 |
| Qy | 6781 | ATTAATGTTGACCGAGAAAAATAGTATCAACTTCTAGCATGATGACCATCGCTTCCATT | 6840 |
| Db | 6781 | ATTAATGTTGACCGAGAAAAATAGTATCAACTTCTAGCATGATGACCATCGCTTCCATT | 6840 |
| Qy | 6841 | GCCTACTGGAATTTCAAAATGAACGAATGCAATGCAATGCAATGCAATGCAATGCAATGCA | 6900 |
| Db | 6841 | GCCTACTGGAATTTCAAAATGAACGAATGCAATGCAATGCAATGCAATGCAATGCAATGCA | 6900 |
| Qy | 6901 | TCCTTACTAGAGTGTATGTTTCTGTTTTCGCTTTGTTTTCGCTTTTTCGCTTTTTCGCTTT | 6960 |
| Db | 6901 | TCCTTACTAGAGTGTATGTTTCTGTTTTCGCTTTGTTTTCGCTTTTTCGCTTTTTCGCTTT | 6960 |
| Qy | 6961 | ATCATTTGTTTTCGCTTTTTCGCTTTTTCGCTTTTTCGCTTTTTCGCTTTTTCGCTTTT | 6992 |

| | | | |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|------|
| Db | 6961 | ATCATTGTTGAGCAAAACAGCAAAAGAGCTT | 6992 |
| RESULT 2 | AF118389 | 15401 bp | DNA |
| LOCUS | Streptococcus suis Cps2A (cps2A), Cps2B (cps2B), Cps2C (cps2C), Cps2D (cps2D), Cps2E (cps2E), Cps2F (cps2F), Cps2G (cps2G), Cps2H (cps2H), Cps2I (cps2I), and Cps2J (cps2J) genes, complete cds; | | |
| DEFINITION | Cps2K (cps2K) gene, partial cds; and unknown genes. | | |
| ACCESSION | AF118389 | | |
| VERSION | AF118389.1 | GI:4580620 | |
| KEYWORDS | Streptococcus suis | | |
| SOURCE | Streptococcus suis | | |
| ORGANISM | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus. | | |
| REFERENCE | 1 (bases 1 to 15401) | | |
| AUTHORS | Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F., Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A. | | |
| TITLE | Identification and characterization of the cps locus of Streptococcus suis serotype 2: the capsule protects against phagocytosis and is an important virulence factor | | |
| JOURNAL | Infect. Immun. 67 (4), 1750-1756 (1999) | | |
| MEDLINE | 99184998 | | |
| PUBMED | 10085014 | | |
| REFERENCE | 2 (bases 1 to 15401) | | |
| AUTHORS | Smith,H.E., Damman,M., van der Velde,J., Wagenaar,F., Wisselink,H.J., Stockhofe-Zurwieden,N. and Smits,M.A. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (07-JAN-1999) Department of Bacteriology, DLO-Institute for Animal Science and Health, P.O. Box 65, Lelystad 8200 AB, The Netherlands | | |
| FEATURES | Location/Qualifiers | | |
| source | 1. .15401 | | |
| | /organism="Streptococcus suis" | | |
| | /mol_type="genomic DNA" | | |
| | /strain="10" | | |
| | /serotype="2" | | |
| | /db_xref="taxon:1307" | | |
| | <1. .721 | | |
| | /note="Orf22" | | |
| | /codon_start=2 | | |
| | /transl_table=11 | | |
| | /product="unknown" | | |
| | /protein_id="AAD24444.1" | | |
| | /db_xref="GI:4580621" | | |
| | /translation="SLDIDHMEVMEASKSAAGSACSPQAYQAAPGAEINIIIVTIT GGLSGFNARVARDMYIEHPNVIHLIDSLASGMDLHVQINRLISAGLDFPOV VEATTHYREHKLPLVAKVDNLVKNGLSKLVCTVGLNIRWGEASAEGLKLLQ KARGHKSVTAAPFEMKAGYDGGRIYMAHRNAKFPQPSSELVKASFPFAIDEVAT SGLCSFYABEGGLMGYEYKA" | | |
| | /complement (820. .2079) | | |
| | /note="Orf2y" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="unknown" | | |
| | /protein_id="AAD24445.1" | | |
| | /db_xref="GI:4580622" | | |
| | /translation="MKKYQVYIIQILGIBEHFRKRGKLPISRQLREQVHCSDTVQ KAMLEKYQNKIYAVEKSGYVILDRDFQDHTCAQSVRLSRITYEDFRICLKESLIG RENLVNHYHQEGLAELISSVQSLMDYHVYTKDQLVITAGSQALYILTOMETLA KTEILLIENPYSRMIELIRHQGIPYQTIERNLDGIDLESIFQFGKIKFPYIPR LHPGSDTIATKVLAKQADVDVIIEDDYADFDSSHSLPHLYDIDNRVIYIK SFTPLPALRIGALSNPQNRDIFIKHSLIDITNLINQKALSIVDINGMFARNQ HLHHIYHAQWNKIKDCLKELALNIPYRIPKGSVTFQLSKGILSPSIQHMFGKCYFSG QKADFLOIFFEODFADKLEQFVRYNE" | | |
| | 2202. .2936 | | |
| | /note="Orf2x" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="unknown" | | |
| | /protein_id="AAD24446.1" | | |
| CDS | | | |
| | /db_xref="GI:4580623" | | |
| | /translation="MKIIPNAKEVNTNLNENASFYLLSDRSKPVLDALISQDFVKMAA FYKLNKAELEADRWYRITGOAKTYPAQLYDGLMYRMYDRGIDSKENYLRDHV RVATYLGIIHPEFISPHRLDPGSKIGNSLKQIYRPFYIDQEVGDDSLIISLASS BEQVSPQIQRLKVLKILFMEKAGQLKVHSTISKGRGLLSWLAKNNIQLSDIQD FKVDGFEYCTSESTANQLTFIRSIM" | | |
| | 3041. .4486 | | |
| | /gene="cps2A" | | |
| | 3041. .4486 | | |
| | /gene="cps2A" | | |
| | /function="putative regulatory protein" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps2A" | | |
| | /protein_id="AAD24447.1" | | |
| | /db_xref="GI:4580624" | | |
| | /translation="MKRGSRSSKPKLVNFALLGLYSTLCLFLVTMYRYNILDPR YLNIYITLLVGAVLAGLIMWRKARIETALLVFSVITSVGIYGMQSVKFPSTR LNSSTSEYEMSLVPANSDITDVROLTSILAPEDQDNITALLDDISKWESTQAT SPQTSYLTAVQSMANGESQAMVFNFTILENEDQPSKVKKIYSFKYQTVETAT KQVSGDSFNIYISGIDAYGPISVSRSDVNIINTVRATHKILLITTPRDSYVAPADG /QGYDKLTHAGIYGVNASVHTLENFYGIDISNYRLNFISFLQIDLVLGGIDVMDP EFTSLGNHYHFPVGGVHLANSQALGFVRERYSLTGGDNRGKQKVIAALIKMSTP ENLKNYQAILSGLEGIOTDLSLETIMSLVNTQLESQTQFTVESQALTGTGRSDLSY AMPGSQLYMMEINQDSLEQSKAAIQSVLVEK" | | |
| | 4504. .5193 | | |
| | /gene="cps2B" | | |
| | 4504. .5193 | | |
| | /gene="cps2B" | | |
| | /function="putative role in chain length determination" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps2B" | | |
| | /protein_id="AAD24448.1" | | |
| | /db_xref="GI:4580625" | | |
| | /translation="MNNQEVNAIEIDVFLFKTIWRKKFILLITAVLTAGLAFYSSF LVTPOYDSTTRIVYVSONVEAGAGLTNQLQAGTYLAKDYREIILSDVLTVATELN LKESLAKISVSPVDTRIVSISVRDADPNEARIANSLTRFAQVKVETKSDVTT LGEAVPAEPTPTPTKRNILLGLLAGILATGLVLMVEVLDDRVKRPQDIEVWGLTL LGIVPDSKKLK" | | |
| | 5203. .5880 | | |
| | /gene="cps2C" | | |
| | 5203. .5880 | | |
| | /gene="cps2C" | | |
| | /function="putative role in chain length determination/export" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps2C" | | |
| | /protein_id="AAD24449.1" | | |
| | /db_xref="GI:4580626" | | |
| | /translation="MAMLEIARTKREGVNTKEYFNARTNIQSGADIKVGIITSVK SNEGKSTTAASLAIAIYARGSKTVLVDADIRNSVMPGFFKPIITKITGLTDYLAGTTDL SQGLCDTIDPNLTVIESKVPNPTALQSKNFENLATLRRYDYVIVDCPPGLGVI DMAIIAQKCDAMVAVVEAGNVKVCSSLKVKQEQLQETGTPFLGVILNKYDIATBKSEY GNYGKKA" | | |
| | 5919. .6650 | | |
| | /gene="cps2D" | | |
| | 5919. .6650 | | |
| | /gene="cps2D" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps2D" | | |
| | /protein_id="AAD24450.1" | | |
| | /db_xref="GI:4580627" | | |
| | /translation="MIDIHSHIIFGVDDGPKTIEESLSLISEAYRQGVRYIVATSHRR KGMFTEPKIMINFLQKLEAVAVYPERICVGAELYSKDILSKLEKKKVPGLNGS CYLLBFSDTPKWEIQEAVNEMTLGLTPVLAHIEYDALAQSERVEKLIKDGCVT QVNSNVLPALIGERAKFEKTRYPLEQDLVHCVASDMHNLYSRFPFWREAYQLVK KEYGEDRAKALFKKNPLLILKNQV" | | |
| | 6675. .8054 | | |
| | /gene="cps2E" | | |
| | 6675. .8054 | | |
| | /gene="cps2E" | | |

Db 14781 TCAGATGCCGTAATTATGGCATAGTTCGCGCAAGGCTGACTTACTTAGCTTTTATAGAC 14840
Qy 6432 TCAGATGATTTTATTCATTCCGAGTTTCATCAACGTTTACGAGCAATTCGAGAGAG 6491
Db 14841 TCAGATGATTTTATTCATTCCGAGTTTCATCAACGTTTACGAGCAATTCGAGAGAG 14900
Qy 6492 AATGCCCTTGTGGCAGTTGCTGGTTATGATAGGCTAGATGCTTCGGGCAATTCCTTAACA 6551
Db 14901 AATGCCCTTGTGGCAGTTGCTGGTTATGATAGGCTAGATGCTTCGGGCAATTCCTTAACA 14960
Qy 6552 GCAGAGCGCTTCTCAAAATCAGGCTGTTCTGAGCGGCAGGAATGTTGTAAAGACTG 6611
Db 14961 GCAGAGCGCTTCTCAAAATCAGGCTGTTCTGAGCGGCAGGAATGTTGTAAAGACTG 15020
Qy 6612 CTAGAGCGGATGTCATCGCTTTGTGGTGCGCTGTATTAACCTCTATAAAAAAGAACTA 6671
Db 15021 CTAGAGCGGATGTCATCGCTTTGTGGTGCGCTGTATTAACCTCTATAAAAAAGAACTA 15080
Qy 6672 TTTCAAGATTTTCCATTGTAAGGCTGAAGATTCATGAAGATGAATCTTCACTTATCGC 6731
Db 15081 TTTCAAGATTTTCCATTGTAAGGCTGAAGATTCATGAAGATGAATCTTCACTTATCGC 15140
Qy 6732 TTGCTCTATGAGTTAGAAAAGTTGCAATAGTTAAGAGTCTTGTACTATTATGTTGAC 6791
Db 15141 TTGCTCTATGAGTTAGAAAAGTTGCAATAGTTAAGAGTCTTGTACTATTATGTTGAC 15200
Qy 6792 CGAGAAATAGTATCAACAATCTTAGCATGACTGACCATCGCTTCCATGCTACTCGAA 6851
Db 15201 CGAGAAATAGTATCAACAATCTTAGCATGACTGACCATCGCTTCCATGCTACTCGAA 15260
Qy 6852 TTTCAAATGAACCAATGCACTTCTCAAGTAGAGGATGAAGAGCTTCTACTAGAG 6911
Db 15261 TTTCAAATGAACCAATGCACTTCTCAAGTAGAGGATGAAGAGCTTCTACTAGAG 15320
Qy 6912 TGTATCGTTCATTTTAGCCTTTGCTGTTGTTGTTTGTAGCAATATATCATGCTGTTG 6971
Db 15321 TGTATCGTTCATTTTAGCCTTTGCTGTTGTTGTTTGTAGCAATATATCATGCTGTTG 15380
Qy 6972 AGCAAAACAGCAAAAGAGCTT 6992
Db 15381 AGCAAAACAGCAAAAGAGCTT 15401

RESULT 3
AY057915 17468 bp DNA linear BCT 01-NOV-2002
LOCUS Streptococcus thermophilus eps type III operon, partial sequence.
DEFINITION
ACCESSION AY057915
VERSION AY057915.1 GI:24473733

KEYWORDS Streptococcus thermophilus
SOURCE Streptococcus thermophilus
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

REFERENCE 1 (bases 1 to 17468)
Rallu, P., Ehrlich, D.S. and Renault, P.
TITLE Diversity of eps operons in Streptococcus thermophilus
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 17468)
Rallu, P., Ehrlich, D.S. and Renault, P.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Genetique Microbienne, INRA, Domaine de

FEATURES
Vilvert, Jouy en Josas 78352, France
source
1. .17468
/organism="Streptococcus thermophilus"
/mol_type="genomic DNA"

Db_xref="taxon:1308"
<1. .17075
/gene="eps type III operon"

<1. .492
/gene="deod"

<1. .492
CDS

gene
CDS
/gene="deod"
/codon_start=1
/transl_table=11
/product="Deod"
/protein_id="AAL23725.1"
/db_xref="GI:24473734"
/translation="RELIVDYGVKLIRVCTAGSLNEDVHVRELVLAAQAATNSNIIR
NWPQYDFQIIASFNLLDKAYHAKNFMGTHVGNVLSDDVFSYNSYFEKNHLELKGWV
KAVEMFAALYYLAAHQVDALAIMTIISDLNPDDEDTTAEERQNTFTDMKVGLETL
IAD"

828. .2288
/gene="eps3A"
828. .2288
/gene="eps3A"
/codon_start=1
/transl_table=11
/product="Eps3A"
/protein_id="AAL23726.1"
/db_xref="GI:24473735"

/translation="MSSRTRNRKQRTGNRSWGMVNVGLTILYAILALVLLFTMFNYNF
LSFRFRIIITIGLLVLAISIFLQTKKSPVTVTVLVIFSLVSLVGFQFKOMIDI
TNRMQTAFASEVEMSIIVPKESDIKDVSQLTSPQPTKVDKXNIILMSALKOKKV
DKVMDVASVQRAYDNLKSGSKAMVLSGSYASLESVDSNFASNLKTIITYIKIKKK
NNSANOVDSKVENIYISGIDTYGSIISTVSRSDVNIIMTVNMNTHKLLTTPRDAYVK
IPGGANQVDKLTAGIYGVETSEOTLEDLYGTIDYARINFTSLKLLDQLGQVTV
HNDQATSLHGKDFPVGDIQMNSEQALGFVRERYSLDGGDNRGRNQEKVISAIVNK
LASLKSVSNTFSIVNNLQDSVQTNISLDTINALANTQLDLSGSKPTVTTSQAVTGTGSG
QUTSYAMPNSISLYMMKLLDNSSVASASQAIKNLMEEK"
2289. .3020
/gene="eps3B"
2289. .3020
/gene="eps3B"
/codon_start=1
/transl_table=11
/product="Eps3B"
/protein_id="AAL23727.1"
/db_xref="GI:24473736"

/translation="MIDVHSHIVFDVDDGPKTLEESLDLIGESYAQGVKIVSTSHRR
KGMFETPKDIFANFSKVAEALYPLDTIYGGELYTLDIVKEKRLNLIIMMHN
QFALIEFSAKTSWKEIHSGLSNVLRAGVTPIVAHIERYDALEENADRVRRIINMGCVT
QVNSHVLKPLFGDKKVRKVRPFLEKNLVHMVASDMHNLGPRPPFMKDAYEIVK
KNYGSKRKNLFIENPKLTLENQYL"
3029. .3721
/gene="eps3C"
3029. .3721
/gene="eps3C"
/codon_start=1
/transl_table=11
/product="Eps3C"
/protein_id="AAL23728.1"
/db_xref="GI:24473737"

/translation="MNQDNTKSDSIDVLALLHLKTKLLILFTAFYFAAFSLGTTF
FIQPTVSTTRIYVNVQDNNLSAQDLQAGTYLANDYKEIITSNDLVSEVTKDEKL
NLSEALSKWVSNIPTDRLISISVNAKTQDAQTLANKRVASAKIKKVTKVEDV
TWLEAKLPESPSPNKLNLVGLGAVLVGVLVREILDORVRPDEDALGM
ALLGIVPDTDKI"
3731. .4489
/gene="eps3D"
3731. .4489
/gene="eps3D"
/codon_start=1
/transl_table=11
/product="eps3D"
/protein_id="AAL23729.1"
/db_xref="GI:24473738"

/translation="MPLLKLVKSNFAKOTERYYNAIRNIOFSGAQIKVIAISSVE
AGEKSTSLNLAISFASVGLTLLIDAUTNSVFSCTFKSNRPYKLSNPLSGNADL
NETICQDTSGLDIVIASGPPVPNPPTSLQNDNFRHLMNEVARSYDYIIDTPPYGLVLI
DAVIAHQADASLLVTAAGKIKRRFVTKAVEQLEQSGSQFLGVNLNKVDMDTVDKYGS
GSYGSYGSYGSYGEYRKTQDTEGSHRHRKRG"
4546. .5913
/gene="eps3E"
4546. .5913


```

/gene="eps3E"
/note="undecaprenylphosphate glucosyltransferase"
/codon_start=1
/transl_table=11
/product="Eps3E"
/protein_id="AAL23730.1"
/db_xref="GI:24473739"
/translation="MKKQERIRIEIGIIQLVWIFAMVASKIPYITITQGSIVLLG
VHVVSYYISYIENKRGXDELIAIVKCYFIPALIAITFLSPADGFSISRGLL
YVWISGVYITVNTLVKYPSSYIRKSNKNIILLSDQARLDNLVRMKDNMDGRI
SAVCVDNPFYDPIKSVKPENLIEYATHSVVDQVLNLPSEQVKIWDYASPFELMG
IPVNLNALFMSQGEKHIQOLGPFKVVTFSTHYSYGDILAKRFLDLCALVGLVL
CGIYFLYPLIRKDGGAIPFAQVRVNGRIFKFKPERSMRVDAEETKKNLMAQNM
SGCMFMDNDRIRIKHFIRKTSDELDPQFNVLKGMSLVGRPTVDEYKYTPE
QKRFLSPGKITGLWQVSGRSEITDFDEVVKLDVAYIDGWTIWRDIOILLKTIKVVVM
KDGAK"
/gene="eps3F"
/note="orf3A"
/pseudo
5946..6264
/gene="orf3A"
/pseudo
5946..6264
/note="orf3A"
/putative rhamnosyltransferase"
/pseudo
6275..7647
/transl_table=11
/codon_start=1
/gene="eps3F"
/pseudo
6275..7647
/gene="eps3F"
/note="Eps3F; putative undecaprenylphosphate
glycosyltransferase"
/pseudo
6275..7647
/transl_table=11
/codon_start=1
/gene="eps3G"
/pseudo
7682..8131
/gene="eps3G"
/pseudo
7682..8131
/note="putative galactosyltransferase enhancer"
/codon_start=1
/transl_table=11
/product="Eps3G"
/protein_id="AAL23731.1"
/db_xref="GI:24473740"
/translation="MKILVGGSSGHLAHLNMLKPFKWKDERFWVTFDKEDARSILKD
EKVSCHEPTNRLNKLKNTFLAKVLKKEKPDVILSSGAVAVPEFYICKLFGAKT
YVIEVFDRIKPTVTGKLVYVPTDKPIVQNEEMKKVYPKAINLGSIF"
8131..8637
/gene="eps3H"
/pseudo
8131..8637
/gene="eps3H"
/note="putative galactosyltransferase"
/codon_start=1
/transl_table=11
/product="eps3H"
/protein_id="AAL23732.1"
/db_xref="GI:24473741"
/translation="MIFVTVGTHQPNRLIOEVDHLVETGVIKEEVFIQYGSYIEP
KFCQWSRLISFDQSKFQKADIIITHGGPATFMSAITNGKKPIVWPQKFGPBHND
HQVDFARNVAKRMGTIIILVEDVDKLGETIENYSIVAELEYDMGNMNNKFKNLEGIV
VKLLDNVK"
8642..9358
/gene="eps3I"
/pseudo
8642..9358
/gene="eps3I"
/codon_start=1
/transl_table=11
/product="Eps3I"
/protein_id="AAL23733.1"
/db_xref="GI:24473742"
/translation="MNKATKCFLLIAHNNWGLKKLIECLDSQTHDIPVHVDRKSKDF
DESYFIDSVTKSLKFRFPTVYMGDFSQVQWEMFLLEQYAVIEGYDYYHIIISGDLPL

```

```

KNNKEIDLFEENKGEKFIILYDNNLTEDNLEISRRTKYHYFLQNYRRRYTENWKNKSPF
TFCBRVSLVQLIIFGVNRKRLDQIKYGSQWVSITNELVKTILHEEBEKITSIFQYTK
CKNHRDSIFISFNSQVSN"
9374..10474
/gene="eps3J"
9374..10474
/gene="eps3J"
/codon_start=1
/transl_table=11
/product="Eps3J"
/protein_id="AAL23734.1"
/db_xref="GI:24473743"
/translation="WKLIQIPGGLFSDGINSFIVEVVTAMDKSEMDIRVLATNNAEE
STLQRVBSGCEVSVIPYRKQNIYKYPFKLRYIAKEKIDIVHHGSSAISMSVELLAA
KLAGCKVRIASHSNTICENRVDQILRRPFDKLYTECFSGOEGKWLFGNRNFIVIP
NGRSLKRYENKQRLQYRKGLHTPDALVIGHVRFNQKNHKLIOIIFYEHHKYP
KSYLVLTGTDVTKVKQOVTEBLENVYIEIGAVNNVPDYLSPMLPSLYEGLP
SWIEMOISGLPCLISDNIINECKITSLVBPESITETPETWAKDIELVELKIEDRNINKD
KIPKEVRDAGYDIDEDAKLKLKELYESLYSPK"
10496..11166
/gene="eps3K"
/pseudo
10496..11166
/gene="eps3K"
/note="Eps3K; putative glucosyltransferase"
/pseudo
Query Match 13.7%; Score 955.6; DB 1; Length 17468;
Best Local Similarity 65.8%; Pred. No. 6.2e-99;
Matches 1484; Conservative 0; Mismatches 729; Indels 41; Gaps 5;
Qy 1 ATCCGCAAAACGAAATGGCATTATTGATATGATAGTAGCAGTGGCAATTCTTCAATCTTAA 60
Db 6294 ACCGCAATTAGACTGGCGCTTTTAGAATTGATGGGTAGTAGTGGCTGTCTCCATTA 6353
Qy 61 CAAGTCATATACCAATGCTGATTTAATCGTCTGGAAATTTTATCATATAATGATGTTTC 120
Db 6354 CCACCCACTTGGCCACCTCAAAGATGAATAGATAGGACGCTATTGATTAATTCGCCGTC 6413
Qy 121 ATTATTTTGCATTTTATATCTCGTATGCGAGTTGAATTTGAGTATAGAGTAACTCTGA 180
Db 6414 ATTTTGGCTTTTATTTCTCAGTATGTCATCAGAAATTTGAGATGAGGCTACTTAA 6473
Qy 181 TAGAGTTTGAATAAACATTTAACTATAGTATAATTTTGAATTTTCTTACGGCAGTAT 240
Db 6474 TAGAGTTTGAACGGTCTAGGTATGTTTGAATTTTCTTACTTCTTCAATTTTCTTTT 6533
Qy 241 CATTTTGTGGAGAAATAATTTCCGACITTTCAAGACGTGGTCCGCTGATTTTCAATTA 300
Db 6534 CTTTTTATTGGGAGATAGTTTATCATTCGCGACGTGGGTGGCATTTATTTACCATTC 6593
Qy 301 TAAACTTCGTTTTCGTATACCTATTAAACGTAATTTATTAAGCAGTTTAAAGGATAGCTTTC 360
Db 6594 TTACCTTATCATCGTTTATTAATTAATCACTGCATGTTTCGTAGATACAAATACCTTCTCT 6653
Qy 361 TATTTTCGACATCTATCAAAAAAGACGATTTAAATTAACACGCTGAAACGATGGGAAA 420
Db 6654 TTATGACTGCTGAACAACAGAGAATACTTTAGTCAATTAACCACTACTCAGAGATTGCGAT 6713
Qy 421 ATATGCAAGTTTATTTGAATCAATAAACAAATTCAAAAAAATCTTGTTCATTTGGTAG 480
Db 6714 TAATGGAGGGGCTCTTTGAATCGGATCAACTTCCTTACCATAATTTTGGCGGTATTGTAG 6773
Qy 481 TTTTGGGTACAGAAATAGATAAATAATTAATTTATCATTCACCTCTATTTATCTGTGGAG 540
Db 6774 TCATTGG---AGACGGGTATGTGGCCTTTTCAGAGGGGAGTTCCAGTAGTGTCTTTTGA 6830
Qy 541 AAGCTATAGAGTTTTCACAAAGGAAGTGGTCGACCACTCTTTTATAAATCTTACCAAGTG 600
Db 6831 ATGCAATAGAGTTTTCGACCCCATGAGTTGTCGACCATGTGTTTATCAACTTACCGAGTG 6890
Qy 601 AGTTTATAGCGTAAAGCAATTCGTTTCAGATTTTGGTGTGTTAGTATTGATGTAAGCG 660

```



```
Db 6891 AACATTACGATCTCAAAACATCTTGTTCCGATTTTGAAGTCATGGGTATTGATGAGTG 6950
Qy 661 TTGATATTAATTCATTCCGTTTACTCGTTTGAAAGACAAAGAAATCCAACCTGCTAGTG 720
Db 6951 TAGATATTAACACATCTGATTTTAGGCGTTTAAACAAATAAAGAAATCAAAACAGGTCGAG 7010
Qy 721 ACCATAGCATTTGAATTTTCCACAAATTTTATAAGCCCTAGTCATATCATGATGAAC 780
Db 7011 ACCATAGCAUGTGACCTTTAACTCCATTAATCTACAAACATAGCCATATCTTCTAAGC 7070
Qy 781 GACTTTTGGATPACTCGGACGGTAGTCGGGTTAAATTAATTTGGGTATAGTTTCTATTT 840
Db 7071 GCATGTTGGATATCTTTGGGCGTTTGATTTGCTTCTAAATTTTGGTCTGTCGGGATTG 7130
Qy 841 TGTTAGTTCGAATTAATTCGTAGACATGGTGACCGGCTATTTTGGCTCAGAAACGAGTTG 900
Db 7131 TCTTAGCCCCCATCATTCGGAAGATGGAGACCGGCCATTTTGTTCAGAAACGAGTAG 7190
Qy 901 GACGAATGGACGCATATTTACATTTCTACAAGTTTTCGATCGATGATGTTGATGCTCAGG 960
Db 7191 GGAAGAACGGACGATCTTTAAATTTTATAATTTCCGTTCTAUGTACACTGATCGGGAG 7250
Qy 961 AGCCCAAAAAAGACTGCTCAGCCAAACCCAGATGCAAGGGTGGGTATGTTTAAAAATGG 1020
Db 7251 A---AAGAAAGAAATGAAGGCACAAATCAGATGCA-----CGTGGAAATGTTCAAGAT 7301
Qy 1021 GAAAAACGATCCTAGAAATTTACTCAATTTGGACATTTTCATAGCGAAAAACAAAGTTTAGAC 1080
Db 7302 GAATAACGATCCACGATATCACTCTATCGGTCGTTTATTTCG-AAAAACGAGTCTGGACG 7360
Qy 1081 AGTTACACACAGTTTATAATGTTTTAAATGGCGGATATGAGTCTAGTTGGTACAGCTCCAC 1140
Db 7361 AACTGCCTCAGTTTACATATGTTGGTCGGCGATATGAGTCTTGTGGGACTCGCCAC 7420
Qy 1141 CTACAGTTGATGAATTTGAAAAATATATCTCTCGTCAAAAGAGACGATGAGTTTAAAC 1200
Db 7421 CAACTGTCGATGAATTTGAAAAATATACCAAGTCAAAACGTCGCTCAGCTTCAAGC 7480
Qy 1201 CAGGATATACAGGTCCTGGCAGGTTAGTGTCTAGTATATATACAGACTTCGACGACG 1260
Db 7481 CAGGAATTTACAGGTCATATGGCAGGCAAGTGGTCGTAGTATATATACCAACTTTGATGAG 7540
Qy 1261 TAGTTCGTTGGACTTAGCATACATGATATTCGACTATCTGCTCAGATATTAATAATTT 1320
Db 7541 TGGTAAGACTCGATGTTGAATATATCGAATACTGTCGATTTGGTCGATATCAAGATAT 7600
Qy 1321 TATTAAAGACAGTGAAGTTGTATTTGTGAGAGGGAAGTAAGT----- 1365
Db 7601 TGTTGAAGACTATTGTCGTGCTTTTAAAGAGAGGGAAGTAGGTAGAGCCCTCTTCA 7660
Qy 1366 -----AAAAGTATATGAAGTTTGTGTTGTCGTTCTTCAGGGGACATTTG 1412
Db 7661 TTCTTGTTATGGGAGAAAAATGAAATTTGTTTAGTAGGGTCATCTGGAGGGCAATTA 7720
Qy 1413 ACTCACTTGATTTGTTTAAACCGTTTGGAGGAAGAAAGCTGTTTGGTGAACATTT 1472
Db 7721 GCACATTTAAATATGTTTAAACCAATCTCGAAGATGAAGAACGTTTTTGGSTHACCTTT 7780
Qy 1473 GATAAAGAGGATCAAGAAGTCTTTTGAAGATGAAAAATGTATCCATGTTACTTTCCA 1532
Db 7781 GATAAAGAGGATCAAGAAGCATCTTAAAGATGAAAAATGTTTCTTGTCACTTTCTCT 7840
Qy 1533 ACAATCGCAATCTCATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGT 1592
Db 7841 ACAATCGAACTTTAAAAATTTAATTAATAATACTTTTCTAGCTATTAAAGTATTAATA 7900
Qy 1593 GATGAGAAACAGATGTTTATTTATCTCGTCGCGCGTGTGTCCTCCCTCTTTTAC 1652
Db 7901 AAAGAAACCTGAATCATTTATTTCACTCGAGCTCGGGAGCCGTTCCATTTCTTTTAC 7960
Qy 1653 ATCGGAAACATTTTGGAGCAAGACGATTTTATATTTGAAGTATTTGATCGAGTTTAAATA 1712
Db 7961 ATTGGTAAACATTTTGGAGCTAAGACGGTTTATATTTGAAGTCTTTTGTAGATGAATTTGATA 8020
```

```
Qy 1713 TCTACATTAACTCGAAAACTAGTTTATCCGTTAACAGATATTTTATGTTTCAGTGGAA 1772
Db 8021 CCAACAGTTTACAGAAAAATTAGTGTACCTCTGTACCGGATAAGTTTCATTGTTCAATGGAA 8080
Qy 1773 GAAATGAAGAAGCTATATCCTAAATCTATTAACTTGGGGAGTATTTTTTAATGATTTTTG 1832
Db 8081 GAGATGAAAAAGTTTATCCAAAGCTATTAAATTAGGAGTATCTTTTAATGATTTTTG 8140
Qy 1833 TAA CAGTAGGAACCTCATGAAACAA CAGTTTAAATCGATTGATAAAGAGATTGATTTATGA 1892
Db 8141 TAA CAGTAGGAACCA CACGAACAA CCTTTTAAATCGCTTAAATTCAGGAAGTGGATCATTTAG 8200
Qy 1893 AAAAAAATGGAAGTATAACCGACGAATAATTTTAAATCAACAGGATATTTCTGACTATATTC 1952
Db 8201 TTGAAA CAGGAGTGA TAAAGAGAGGTC TTTTATCAACAGGATATTTCTATCTACGAAC 8260
Qy 1953 CAGAAATATTGCAAGTATAAAAAATTTCTCAGTTTACAAAGAAATGGAAACAATATATTAAACA 2012
Db 8261 CTAAATTTTGTCAAGTGTCTCGTTTGTATAGTTTGTGATCAATGAGTAATTTTATGCAAA 8320
Qy 2013 AATCAGAAGTAGTTATTTGCCACGAGGCCCTACTTTTATGAAATTCATTTATCCAAAG 2072
Db 8321 AAGCTGATATTATATCACTCACCGAGGTC CAGCGACTTTTATGAGTGCCATTACAATG 8380
Qy 2073 GAAAAAACAATTTTGTTCCTAGCAAAAAAAGTATGCTGAACATGTAATGATCATC 2132
Db 8381 GCAAAAAGCCAA TTTGTTGTCCTAGACAAAGAAAGTTTGGTGAACATGTCAATGACCATC 8440
Qy 2133 AAGTAGAGTTTGAAGAAGAAATTTTACAGATAATAATTTTATTTATAGAAAAATATAG 2192
Db 8441 AAGTAGATTTTGCAGAAATGTAGCAAAAGATGGAACTATCATTTCTGTTGAAGATG 8500
Qy 2193 ATGATTTGTTGAAAAAATTAATTTGAAAGTTTCTAA 2226
Db 8501 TAGATAAATTTGGGGAGACTATTGAAAAATTACAA 8534

RESULT 4
AX283715 LOCUS AX283715 17468 bp DNA linear PAT 20-NOV-2001
DEFINITION Sequence 9 from Patent WO0179500.
ACCESSION AX283715
VERSION AX283715.1 GI:17044426
KEYWORDS Streptococcus thermophilus
SOURCE Streptococcus thermophilus
ORGANISM Streptococcus thermophilus
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Rallu, F., Besancon-Yoshpe, I., Fremaux, C., Mengaud, J. and Renault, P.
TITLE Streptococcus thermophilus operons involved in exopolysaccharide (e
ps) synthesis
JOURNAL Patent: WO 0179500-A 9 25-OCT-2001;
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE (INRA) (FR) ;
COMPAGNIE GERVAILIS-DANONE (FR) ; Rhodia Chimie (FR)
FEATURES
source 1.17468
location/Qualifiers
/organism="Streptococcus thermophilus"
/mol_type="unassigned DNA"
/db_xref="taxon:1308"
ORIGIN
Query Match 13.7%; Score 955.6; DB 6; Length 17468;
Best Local Similarity 65.8%; Pred. No. 6.2e-99;
Matches 1484; Conservative 0; Mismatches 729; Indels 41; Gaps 5;
Qy 1 ATCGCCAAACGAAATTTGGCATTTATTGATATGATAGTAGCAGTTGCAATTTCTGCAATCTTAA 60
Db 6294 ACCGGCAATTTTAGCTGGCGCTTTTGAATTTGATTTGGGTAGTAGTGCTGTCTCCATTA 6353
Qy 61 CAAGTCATATACCAAAATGCTGATTTAAATCGTTCTGGAATTTTATCAATGATGCTTC 120
```


Db 6354 CCACCCACTTGCACACCTCAAGATGAATCAGATAGGACGCTATTGATTATTGCCCTGC 6413
Qy 121 ATATATTTTGCATTTTTTATATCTCGTATGCGAGTTGAAATTTGAGTATAGAGTAACTCTGA 180
Db 6414 ATTTTGTGGCTTTTATTTCTCAGTATGTCTATCAGATATTTGAGAGTAGAGGCTACTTAA 6473
Qy 181 TAGAGTTTGAAGAAACATTTAACTATAGTATTAATTTTGCAATTTTCTTACGGCAGTAT 240
Db 6474 TAGAGTTTGAACGGGCTCGCTAGGTATTTGATTTTCCGATTTATCTCTCATTTTCATTTT 6533
Qy 241 CATTTTCTGTCGAGAAATTAATTTGCGACATTTCAAGACGTTGGTCCGCTGATTTTACATTAA 300
Db 6534 CTTTTATTTGCGAGATAGTTTATCATATGCGACGTTGGTGGCATATTTTACCAATC 6593
Qy 301 TAAACTCTGTTTGGTATACCTATTATTAACGTAAATTTAAGCAGTTTAAAGATAGCTTTC 360
Db 6594 TTACTTTATCATCTGTTTATATATATCACTGCACCTGTTTCGTAGATACAAATACCTCTCT 6653
Qy 361 TATTTTCGACAAATCTATCAAAAAGACGATTTCTAATTTACAAAGGCTGAAAGTGGGAAA 420
Db 6654 TTATGACTGCTGAAACACAGAGAATATCTTTAGTCAATTTACCACACTACTGAGAGATTGCCAT 6713
Qy 421 ATATGCAAGTTTATTTGAATCACAATAACAAATTTCAAAAATCTTGTTCGATTTGGTAG 480
Db 6714 TAATGGAGGGCTCTTTGAATCGGATCAACTTCTTACCAAAATATTTGGCGGGTATTTGAG 6773
Qy 481 TTTTAGTACAGAAATAGATAAAATTAATTTATCAATACCGCTCTATTTCTGTGGAAG 540
Db 6774 TCATTTGG---AGACGGTATGTGGCTTTCCAGAGGAGTTCCAGTAGTCTCTTTTGATT 6830
Qy 541 AAGCTATAGATTTTCAACAGAGGAAGTGTGCGACACGCTTTTATATAAATCTACCAAGTG 600
Db 6831 ATGCAATAGATTTGCGAGCCCATGAAGTTGTGCGACCATGTGTTTATCAACTTACCGAGTG 6890
Qy 601 AGTTTATAGCGTAAGCAATTCGTTTTCAGATTTTGTAGTTTGTAGTATTCATGTAAGC 660
Db 6891 AACATTACGATCTCAACATCTGTTTCCGATTTTGAAGTCATGGGTATTCATGTAAGTG 6950
Qy 661 TTGATATTAATTCATTCGGTTTCTAGCGTTTGAAGAAACAAAATAACAACTGCTAGGTG 720
Db 6951 TAGATATTTACACATCTGATTTTAGGGCTTTTAAACAAATTAACAAATCAACAGGTTGGAG 7010
Qy 721 ACCATAGCATTTGTAATTTTCCAAATTTTATAAGCCTAGTCATATCATGATGAAC 780
Db 7011 ACCATAGCATCTGTGACTTTTAACTCCAATTTACTACAAACATAGCCATATCTTCTAAGC 7070
Qy 781 GACTTTTGGATATCTCGAGCGGTAGTCGGGTTAATTTATTTGGTATAGTTCTTATTT 840
Db 7071 GCATGTTTGGATATCTTTGGGGCTTTGATTTGGTCTTCTAAATTTTGTGCTGTCGGGATTG 7130
Qy 841 TGTAGTTTCCAAATTTCTAGAGATGCTGGACCGGCTATTTTGTCTCAGAAACGAGTTG 900
Db 7131 TCTTAGCCCCCATCATTCGGAAGATGAGAGCCGCGCATTTTGTTCAGAAACGAGTAG 7190
Qy 901 GACAGATGAGCGCATATTTTCAATTTCTACAAAGTTTCGATCGATGATGTTGATGCTGAGG 960
Db 7191 GGAAAAACGAGCATCTTTAAATTTTATAAATTTTAAATTTCCGTTCTATGTACACTGATGCGAGG 7250
Qy 961 AGCGCAAAAAAGACTGTCTAGCCCAAAACCGAGTCAAGGGTGGGTATGTTTTAAATGG 1020
Db 7251 A-----AAGAAAGAAATGAAGGCACAAATCAGATGCA-----CGGTGGAATGTTCAAGAT 7301
Qy 1021 GAAAAACCATCTAGAAATTTACTCCAAATTTGGCATTTTCATACGCAAAAAACAAGTTTAGACG 1080
Db 7302 GAATTAACGATCCAGTATCACTCTCATCTCGTTCGTTTATTCG-AAAAACGAGTCTGGACG 7360
Qy 1081 AGTTACCAAGTTTTTATTAATGTTTTAAATTTGGCGATATGAGTCTAGTCTGGTACAGCTCCAC 1140
Db 7361 AACTGCCTCAGTTTTTCAATATGTTTGGTCTGGCGATATGAGTCTTGTGGACTCGCCAC 7420
Qy 1141 CTACAGTTGATGAATTTGAAAAATATCTCTCGGTCAAAAGAGACGATTTGAGTTTAAAC 1200
Db 7421 CAACCTGTCGATGAATTTGAAAAATATACCAAGTCAAAACGTCGCGCTCAGCTTCAAGC 7480

Qy 1201 CAGGATTAACAGGTCTCTGGCAGGTTAGTGGTCTAGTAATATCACAGACTTTCGACGAGC 1260
Db 7481 CAGGAATTAACAGGTCTATGGCAGGCAAGTGGTCTAGTAATATTAACAACTTTCATGAGG 7540
Qy 1261 TAGTTCGGTTGGACTTACATATGATTAATTCGACTATCTGTCAGATATTAATAATTT 1320
Db 7541 TGGTAAGACTCGATGTTGGAATATATCGATAACTCGTTCGATTTGGTCGGATATCAAGATAT 7600
Qy 1321 TATTAAGACAGCTGAAAGTTCGTATTTGTGAGAGAGGAAAGTAAAGT----- 1365
Db 7601 TGTTCAGACTTATTTGTCGTGTTTTTAAAAAGAGGGAAGTAGTAGAGCCCCCTCTTCA 7660
Qy 1366 -----AAAAAGTATATGAAAGTTCGTTTTGTTGGTTCGTTCTTACAGGGGACATTTG 1412
Db 7661 TTTCTGTTATTTGGGAGAAAAAATGAAATTTGTTTAGTAGGCTCATCTGAGGGCATTTA 7720
Qy 1413 ACTCAGCTGTTATTTTAAACCGTTTTTGGNAGGAAGAACGTTTTTGGGTAAACATTT 1472
Db 7721 GCACATTTAAATATGTTTAAACCAATCTCGGAAGATGAAGAACGTTTTTGGGTAAACCTTT 7780
Qy 1473 GATAAAGAGGATCAAGAAAGTCTTTTGAAGATGAAGAAATGATCCATGTTACTTTCCA 1532
Db 7781 GATAAAGAGGATCAAGAGCATCTTAAAGATGAAGAAATGATTTCTTGTCATTTCTT 7840
Qy 1533 ACAATCGCAATCTCATTAATTTAGTGAAGAAATACTTTTCTTAGCTTTCAAAATTTTACGT 1592
Db 7841 ACAATCGAACTTTAAAAAATTTAATTTAAAAAATACTTTTCTAGCTATTTAAAGTATTA 7900
Qy 1593 GATGAGAAACAGATGTTATTTTTCATCTGGTGGCGGCTGCTGTCCTCTTCTTTTAC 1652
Db 7901 AAAGAAAAACCTGATGTCATTTATTTCTCTGGAGCTCGGTAGCCGTTCCATTTCTTTTAC 7960
Qy 1653 ATCGGAAAACTATTTGGAGCAAGACGATTTATATGAAGTATTTGATCGAGTTTAATAA 1712
Db 7961 ATTGTAACCTATTTTGGAGCTAAGCGTTTATATTTGAAGTCTTTTATGAGATTTGATTA 8020
Qy 1713 TCTACATTAACCTGGAAAACTAGTTTATCCGTAACAGATATTTTATTTGTTTCAAGTGGAA 1772
Db 8021 CCAACAGTTACAGAAAAATTAGTGTACCTCTGACGGATAGTTTCATTTGTTCAATGGGAA 8080
Qy 1773 GAAATGAAGAGGTATATCTTAAATCTATTAACCTTGGGAGTATTTTAAATGATTTTGG 1832
Db 8081 GAGATGAAAAAGTTTATCCAAAAAGCTATTAATTTAGGAGTATCTTTTAAATGATTTTGG 8140
Qy 1833 TAAAGTAGGAACATCATGCAACAGTTTAAATCCATTCATATAAGAGATTTGATTTATGA 1892
Db 8141 TAAAGTAGGAACACACGAAACACCTTTTAAATCGCTTAAATTCAGGAAGTGGATCAATTTAG 8200
Qy 1893 AAAAAAATGGAAGTATTAACCGACGAAATATTTTATTTCAAAACAGGATATTTCTGACTATATTC 1952
Db 8201 TTGAAACAGGAGTATTAAGAGAGAGGTCTTTTATTTCAAAACAGGATATTTCTATCTACGAC 8260
Qy 1953 CAGAAATATGCAAGTATAAAAAATTTCTCAGTTTCAAGAAATGGAACAATATATTAACA 2012
Db 8261 CTAAATTTTGTTCAGTGTCTCGTTTGTATGAGTTTGTATGATCAAAATGAGTAAATTTTATGCAA 8320
Qy 2013 AATCAGAGTATGTTTATTTGCCACGAGGCCGCTACTTTTATGAAATTCATTTCAATTCCAAAG 2072
Db 8321 AAGCTGATATTTATTTATCACTCACGGAGGTCAGCGACTTTTATGAGTGCCATTTACAAATG 8380
Qy 2073 GAAAAACAAATTTATTTCTTAGCAAAAAAGTATGTTGAACATCTAAATGATCATC 2132
Db 8381 GCAAAAAGCCAAATTTGTTGCTAGCAAGAAAAAGTTTGGTGAACATGCTCAATGACCATC 8440
Qy 2133 AAGTAGAGTTTGTAAAGAGAAATTTTACAGATAATAATTTATTTTATAGAAAAATATAG 2192
Db 8441 AAGTAGATTTTGCAGAAATGCTAGCAAAAAGAAAGTGGGAACATCTCATTTCTGTTGAAGATG 8500
Qy 2193 ATGATTTGTTGAAAAAATTTATTTCAAGTTTCTAA 2226
Db 8501 TAGATAAATTTGGGGAGACTATTTGAAAAATTAACA 8534

| | | | | | |
|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|-----|--------|-----------------|
| RESULT 5 | STH289861 | 21562 bp | DNA | linear | BCT 04-JAN-2002 |
| LOCUS | Streptococcus thermophilus eps locus. | | | | |
| DEFINITION | Streptococcus thermophilus eps locus. | | | | |
| ACCESSION | AJ289861 | | | | |
| VERSION | AJ289861.1 | GI:18076389 | | | |
| KEYWORDS | deod gene; Deod protein; Eps 2; Eps 3; Eps 4; eps10 gene; Eps10 protein; eps11 gene; Eps11 protein; eps12 gene; Eps12 protein; eps14 gene; Eps14 protein; eps15 gene; Eps15 protein; eps16 gene; Eps16 protein; eps2 gene; Eps2 protein; eps3 gene; Eps3 protein; eps4 gene; Eps4 protein; eps5 gene; Eps5 protein; eps9 gene; Eps9 protein; epsA gene; EpsA protein; epsB gene; EpsB protein; epsC gene; EpsC protein; epsD gene; EpsD protein; epsE gene; EpsE protein; ORF1; ORF6; pseudogene. | | | | |
| SOURCE | Streptococcus thermophilus | | | | |
| ORGANISM | Streptococcus thermophilus | | | | |
| REFERENCE | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus. | | | | |
| | 1 | | | | |
| AUTHORS | Pluvinet,A., Charron-Bourgoin,F., Morel,C. and Decaris,B. | | | | |
| TITLE | Implication of horizontal transfers in the chimeric structure of the eps locus of Streptococcus thermophilus IP6757 | | | | |
| JOURNAL | Unpublished | | | | |
| REFERENCE | 2 (bases 1 to 21562) | | | | |
| AUTHORS | Pluvinet,A. | | | | |
| TITLE | Direct Submission | | | | |
| JOURNAL | Submitted (18-MAY-2000) Pluvinet A., Laboratoire de Genetique et Microbiologie, Universite Henri Poincare - Faculte des Sciences, B.P. 239, Vandoeuvre-les-Nancy 54506, FRANCE | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1..21562 | | | | |
| CDS | /organism="Streptococcus thermophilus" | | | | |
| | /mol_type="genomic DNA" | | | | |
| | /db_xref="taxon:1308" | | | | |
| | <1..465 | | | | |
| | /note="ORF1" | | | | |
| | /codon_start=1 | | | | |
| | /transl_table=11 | | | | |
| | /product="hypothetical protein" | | | | |
| | /protein_id="CAC82000.1" | | | | |
| | /db_xref="GI:18076390" | | | | |
| gene | /db_xref="UniProt/TREMBL:Q8VM12" | | | | |
| | /translation="DIIVGVCLGAQLLSVAYGKYEHSPEIREIGVPLTLTDEGLAD EHKDFGSTLNTGHWGMFGLSDKAVLATSGYPHQIICFSPKHVAFQAHLEFDLE AIDLIIAADGEEHLYQONKQLDFVQTPEQLNHDYSQMNKKLFAFLDSLTFQI" | | | | |
| | 800..1809 | | | | |
| | /gene="deod" | | | | |
| | 800..1510 | | | | |
| | /gene="deod" | | | | |
| | /codon_start=1 | | | | |
| | /evidence=not_experimental | | | | |
| | /transl_table=11 | | | | |
| | /product="Deod protein" | | | | |
| CDS | /protein_id="CAC82001.1" | | | | |
| | /db_xref="GI:18076391" | | | | |
| | /db_xref="GOA:Q8VM11" | | | | |
| | /db_xref="UniProt/TREMBL:Q8VM11" | | | | |
| | /translation="MSIHSAKQGDIAKILLPGDPLRAKFAIENFLEDVACFNEVRN MFGTYGVYKGRISVIRISMGFMPSISYARELIVDYGKKLIRVGTAGSLNENHVRE LVLAQAATNSIIRNDPOYDFPQIANFNLLDKAVHIAKNFGMTTHGVNLSDDVYF SNYFEKNIELGKWGVKAVEAMEAALYYLAQAQHQVDALAIMTISDSLVNPDYTAER QNTFTDMKVGLETLIAD" | | | | |
| | 1744..1764 | | | | |
| | /gene="deod" | | | | |
| | 1781..1786 | | | | |
| | /gene="deod" | | | | |
| | 1804..1809 | | | | |
| terminator | /gene="deod" | | | | |
| | 1853..3307 | | | | |
| | /gene="deod" | | | | |
| | 1853..3307 | | | | |
| | /gene="epsA" | | | | |
| | 1853..3307 | | | | |
| | /gene="epsA" | | | | |
| | /function="putative exopolysaccharide regulation" | | | | |
| | -35_signal | | | | |
| | -10_signal | | | | |
| gene | /gene="deod" | | | | |
| | 1853..3307 | | | | |
| | /gene="epsA" | | | | |
| | 1853..3307 | | | | |
| | /gene="epsA" | | | | |
| | 1853..3307 | | | | |
| | /gene="epsA" | | | | |
| | 1853..3307 | | | | |
| | /gene="epsA" | | | | |
| | 1853..3307 | | | | |
| CDS | /gene="epsB" | | | | |
| | 3308..4039 | | | | |
| | /gene="epsB" | | | | |
| | 3308..4039 | | | | |
| | /gene="epsB" | | | | |
| | 3308..4039 | | | | |
| | /gene="epsB" | | | | |
| | 3308..4039 | | | | |
| | /gene="epsB" | | | | |
| | 3308..4039 | | | | |
| gene | /gene="epsC" | | | | |
| | 4048..4740 | | | | |
| | /gene="epsC" | | | | |
| | 4048..4740 | | | | |
| | /gene="epsC" | | | | |
| | 4048..4740 | | | | |
| | /gene="epsC" | | | | |
| | 4048..4740 | | | | |
| | /gene="epsC" | | | | |
| | 4048..4740 | | | | |
| CDS | /gene="epsC" | | | | |
| | 4048..4740 | | | | |
| | /gene="epsC" | | | | |
| | 4048..4740 | | | | |
| | /gene="epsC" | | | | |
| | 4048..4740 | | | | |
| | /gene="epsC" | | | | |
| | 4048..4740 | | | | |
| | /gene="epsC" | | | | |
| | 4048..4740 | | | | |
| gene | /gene="epsD" | | | | |
| | 4750..5490 | | | | |
| | /gene="epsD" | | | | |
| | 4750..5490 | | | | |
| | /gene="epsD" | | | | |
| | 4750..5490 | | | | |
| | /gene="epsD" | | | | |
| | 4750..5490 | | | | |
| | /gene="epsD" | | | | |
| | 4750..5490 | | | | |
| CDS | /gene="epsD" | | | | |
| | 4750..5490 | | | | |
| | /gene="epsD" | | | | |
| | 4750..5490 | | | | |
| | /gene="epsD" | | | | |
| | 4750..5490 | | | | |
| | /gene="epsD" | | | | |
| | 4750..5490 | | | | |
| | /gene="epsD" | | | | |
| | 4750..5490 | | | | |
| gene | /gene="epsE" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsE" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsE" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsE" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsE" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsE" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsE" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsE" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsE" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsE" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsF" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsF" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsF" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsF" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsF" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsF" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsF" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsF" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsF" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsF" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsG" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsG" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsG" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsG" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsG" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsG" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsG" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsG" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsG" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsG" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsH" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsH" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsH" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsH" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsH" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsH" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsH" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsH" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsH" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsH" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsI" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsI" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsI" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsI" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsI" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsI" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsI" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsI" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsI" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsI" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsJ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsJ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsJ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsJ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsJ" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsJ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsJ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsJ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsJ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsJ" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsK" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsK" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsK" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsK" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsK" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsK" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsK" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsK" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsK" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsK" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsL" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsL" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsL" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsL" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsL" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsL" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsL" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsL" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsL" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsL" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsM" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsM" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsM" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsM" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsM" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsM" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsM" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsM" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsM" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsM" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsN" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsN" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsN" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsN" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsN" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsN" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsN" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsN" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsN" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsN" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsO" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsO" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsO" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsO" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsO" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsO" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsO" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsO" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsO" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsO" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsP" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsP" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsP" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsP" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsP" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsP" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsP" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsP" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsP" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsP" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsQ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsQ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsQ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsQ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsQ" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsQ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsQ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsQ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsQ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsQ" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsR" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsR" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsR" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsR" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsR" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsR" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsR" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsR" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsR" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsR" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsS" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsS" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsS" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsS" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsS" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsS" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsS" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsS" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsS" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsS" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsT" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsT" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsT" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsT" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsT" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsT" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsT" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsT" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsT" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsT" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsU" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsU" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsU" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsU" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsU" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsU" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsU" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsU" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsU" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsU" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsV" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsV" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsV" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsV" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsV" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsV" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsV" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsV" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsV" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsV" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsW" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsW" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsW" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsW" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsW" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsW" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsW" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsW" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsW" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsW" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsX" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsX" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsX" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsX" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsX" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsX" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsX" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsX" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsX" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsX" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsY" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsY" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsY" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsY" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsY" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsY" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsY" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsY" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsY" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsY" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsZ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsZ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsZ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsZ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsZ" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsZ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsZ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsZ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsZ" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsZ" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsAA" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsAA" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsAA" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsAA" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsAA" | | | | |
| | 5547..6914 | | | | |
| CDS | /gene="epsAA" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsAA" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsAA" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsAA" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsAA" | | | | |
| | 5547..6914 | | | | |
| gene | /gene="epsAB" | | | | |
| | 5547..6914 | | | | |
| | /gene="epsAB" | | | | |
| | | | | | |

| | | | | |
|------|----------------------------------|------|------------------------------------------------------------------|------|
| gene | /gene="epsE" | 497 | AGATAAAATTAATTTATCAATTACCGCTCTATTATTCTGTGGAAGAAGCTATAGAGTTTTC | 556 |
| | /pseudo | 7277 | AGTGGCGTTTCCAGAGGAGTCCCAATAATTCCTTTT---GATGATCGATTGACTTTGC | 7333 |
| CDS | /codon_start=1 | 557 | AACAAGGAGTGTGCGACCAAGCTTTTATAAATCTACCAAGTGAAGTTTGTAGACGTAAA | 616 |
| | /evidence=not_experimental | 7334 | GACTCATGAAGTTGTGCAACCATGTGTTTATCAACTTACCGAGTGAACATTTACGATCTCAA | 7393 |
| gene | /transl_table=11 | 617 | GCAATTCGTTTTCAGATTTTGAGTTGTTAGGTATTGATGTAAGCGTTGATATTAATTCATT | 676 |
| | /product="EpsE protein" | 7394 | ACATCTGTTTCCGATTTTGAAGTCATGGGTATTGATGTGAGTGTAGATATTTACTTATT | 7453 |
| CDS | /db_xref="PSEUDO:CAC8206.1" | 677 | CGGTTTACTGCGTTGAAAAACAAAAATCCAACTGCTAGGTGACCAATAGATTTGTAAC | 736 |
| | /db_xref="GI:18076396" | 7454 | TGATTTGAGGCTTTAAAAAATAAAAAAATCAAAACAAGTTGAGAGACCATAGTATCGTGAC | 7513 |
| gene | /gene="eps1" | 737 | TTTTTCCACAAATTTTATAGCCTAGTCATATCATGATGAAGACATTTTGGATATACT | 796 |
| | /pseudo | 7514 | TTTTTAACCTCCAAATTACTACAAACATAGCCATATCTTTCTTAAAGCGCATGTTGGATATCTT | 7573 |
| CDS | /function="glycosyltransferase" | 797 | CGGAGCGTAGTCGCGTTTAATTTATTTGGGTATAGTTTCTATTTTGTAGTTTCCAAATTAT | 856 |
| | /codon_start=1 | 7574 | TGGGCTGTGATTTGGTCTTCTAAATTTGGTCTCGGTGGGATTTGCTTAGCTCCTATCAT | 7633 |
| gene | /evidence=not_experimental | 857 | TCGTAGAGATGGTGG-ACCGGCTATTTTGTCTCAGAAAACGAGTTGGACAGAAATGCAACGA | 915 |
| | /transl_table=11 | 7634 | TCGGAAGATGGAGGAACCTGCTATTTTGGTTTCAGAAACAGTAGTAGGAAAAAACGGTCGTA | 7693 |
| CDS | /product="Eps2 protein" | 916 | TATTTACATCTCACAAGTTTCGATCGATGATGTTGATGCTGAGAGCGGCAAAAAGACT | 975 |
| | /protein_id="CAC82008.1" | 7694 | TCITTAATTTTATAAATTCGTTCTATGTATATAGATGCAAAAAACGGAAAAAAGAAC | 7753 |
| gene | /db_xref="UniProt/TrEMBL:Q8VM08" | 976 | TGCTCAGCAAAACAGATGCAAGGTGGGTATGTTTAAATCGGAAAAACGATCCTAG | 1035 |
| | /db_xref="GI:18076396" | 7754 | TTATGGCTCAAAACAGATGCA-----GGGTGGTATGTTTAAATGGATATATGACCCACG | 7808 |
| CDS | /gene="eps3" | 1036 | AATTACTCCAATTGGACATTTTCATACGCAAAACAAAGTTTAGACAGGTTTACCAGATTTT | 1095 |
| | /function="glycosyltransferase" | 7809 | GATTTACACCAATTTGGTCAGTTTATCCG-AAAAACAAGTCTAGATGAATTTGCCACAGTTTT | 7867 |
| gene | /codon_start=1 | 1096 | ATAATGTTTAAATTTGGCGATATAGTCTAGTTGGTACACGTCACCTTACAGTTGATGAAT | 1155 |
| | /evidence=not_experimental | 7868 | ATAATGTAAGTTTGGCGATATAGGCTTTGGTAGTACTCTCGTCCACCGACAGTGGATGAAT | 7927 |
| CDS | /transl_table=11 | 1156 | TTGAAAAATATACCTCTGCTCAAAGAGACGATTTGAGTTTAAACCCAGGATTCACAGTCT | 1215 |
| | /product="Eps3 protein" | 7928 | TTGAAAAAGTATACACCAAGTCAAAAACGTCGCCCTCAGCTTCAAGCCAGAAATTACAGTCT | 7987 |
| gene | /db_xref="UniProt/TrEMBL:Q8VM07" | 1216 | TCTGCGAGGTTAGTGGTCGTAGTAATATACAGACTTTCGACGAGTGTTCGGTTGGACT | 1275 |
| | /db_xref="GI:18076397" | 7988 | TATGCGCAGGTAAGTGGTCGTAGCAATATTACCAACTTTGATGAGTGTGTAATAAATCTCGATG | 8047 |
| CDS | /gene="eps4" | 1276 | TAGCATCATTTGATAAATTTGGACTATCTGGTCAGATATTTAAATTTTATTAAGACAGTGA | 1335 |
| | /function="glycosyltransferase" | 8048 | TTGAATATATCGATAAATCTGGTCGATTTGGTCGGATATCAAGATATTTGTTGAAGACTATCG | 8107 |
| gene | /codon_start=1 | 1336 | AGTTGTTATTTGAGAGAGGGAAGTAAGTAA--AGTATATGAAGTTTCTGTTGGTCTCG | 1393 |
| | /evidence=not_experimental | 8108 | TCGCGGTTTTTCAAAAAAGAGGGAAGTAAGTAGATTATTTATGAAGTTTCTGTTAGTAGG | 8167 |
| CDS | /transl_table=11 | 1394 | TTCTTCAGGGGACATTTGACTCACTTGTATTGTTTAAACCGTTTTCGGAAGGAAGAAGA | 1453 |
| | /product="Eps4 protein" | 8168 | TTCTTCTGGTGGGCAATTTGGCACACTTGAATATGCTTTAAACCTTTTGGAGTGAACAGGA | 8227 |
| gene | /db_xref="PSEUDO:CAC82010.1" | 1454 | ACGTTTTTGGGTAACTTTTGATAAAGAGGATGCAAGAAGTCTTTTGAAGAATGAAAAAT | 1513 |
| | /db_xref="GI:18076398" | 8228 | CCGCTTTTGGGTTTACCTTTTGATAAAGAAGATGCAAGAAGTATTTTAAAGATGAGCACTT | 8287 |
| CDS | /gene="eps1" | 1514 | GTATCCATGTTTACTTTTCCAAACAATTCGCAATCTCATTTAATTTTAGTGAATAATCTTCTT | 1573 |
| | /pseudo | 8288 | TTATCCGCTGCTATTTTCTCTACTAACAGAAATTTTAAAAAATTTTAGTAAAGAATACTTTTT | 8347 |

Query Match 12.8%; Score 892.6; DB 1; Length 21562;
Best Local Similarity 66.9%; Pred. No. 7e-92;
Matches 1393; Conservative 0; Mismatches 669; Indels 19; Gaps 8;

| | | | |
|----|------|---------------------------------------------------------------|------|
| Qy | 317 | ATACCTATTAAACGPAATTTAAGCAGTTTAAAGATAGCTTCTATTTTCGACAATCTA | 376 |
| Db | 7097 | ATACCTTTAATCACTGCACTGTCGGTAGATACAAATATCTATCTTATGACTGCTGAACA | 7156 |
| Qy | 377 | TCAAAAAAGACGATTTCTAAATTACACCGGCTGAACGATGGGAAAAATATGCAAGTTTATT | 436 |
| Db | 7157 | ACAGNAGATACATTAGTCATTTACCACCACTGAGAGATTCGATTAATGGAGGGCTCTT | 7216 |
| Qy | 437 | TGAATCATATAACAAATTTCAAAAAAATCTTTGTCATTTGGTAGTTTGTAGTACAGAAAT | 496 |
| Db | 7217 | TGAGTCAGATCAACTTCTCTCAATAATTTGGCAGGTATAGTAGTCATCGGAGATGGTGA | 7276 |

Qy 436 TTGAATCACATAAACAATTCAAAAAATCTTGTTCATTTGGTAGTTTATAGGTACAGAAA 495
Db 5455 TCGTAGAATCTGATGAACCTTTTCATGGAACCTGGTAGCAGTAAGTGTGTTGGATAAATCTG 5514
Qy 496 TAGATAAAATTAATTTATCAATACCGCTCTATTTATCTGTGGAAGAAGCTATAGAGTTTT 555
Db 5515 ATTTTCAACATGATAAATAACCTCTGAATGAAA- - - - -GGAATAAATTAATGAATTTG 5568
Qy 556 CAACAGGGAAGTGGTCGACACAGCTCTTTATAAATCTACCAAGTAGGTTTTTACAGCTAA 615
Db 5569 CAACGCATGAAGTTGTGGATGAGGTGTGTGTCGATCTTCCAGGAGAGAGCTATCATATTG 5628
Qy 616 AGCAATTCGTTTCAGATTTTCAGTTGTGTAGTATTTGATGTAAGCGTTGATATTAATTCAT 675
Db 5629 GAGAAATTAATCTCTAAGTTTGAACAATGGGATAGATGTAACTGTAAATCTTAATGCAT 5688
Qy 676 TCGGTTTTTACTCGCTTGAAAAACAAAAATCCAACTGCTAGGTGACCATAGCATTTGTAA 735
Db 5689 TTAATAAGAATTTGGGTGCTAATAAACAATTCATGAGATAGTAGGATTTGAATCTAGTCA 5748
Qy 736 CTTTTTCACAATAATTTTAAAGCTAGTCAATCATCATGATGAACAGACTTTTGGATATAC 795
Db 5749 CTTTCTCTACAAATTTTATAAACTAGTCAATGATGATTTCAAGAGAGAAATCTCGATATT 5808
Qy 796 TCGGAGCGGTAGTCGGGTTAATTTTGTGATAGTTTCTATTTCTAGTTCCCAATTA 855
Db 5809 GTGGTGCCATTAATTTGGCCCTTAATCTTTTGTCTATAGCTAGTCTAGTTTATGTTTCAATTGA 5868
Qy 856 TTCTGATAGATGTTGGACCGGCTATTTTGTCTCAGAAACGAGTTTGGACAGAAATGGACGCA 915
Db 5869 TTCTGTAAGATGGCGGACCACTATTTTGTCTCAACTGTTATAGGGAATAATGGTCGAC 5928
Qy 916 TATTTACATTTCTCAAGTTTCGATCGATGTATGTTGATGCTGAGGAGCGGCAAAAAGACT 975
Db 5929 ATTTTACCTTTTATAAATTCGGTTTCGATCGGATCGATGCTGAAGTTTATCAAGAACAGT 5988
Qy 976 TGCTCAGCAGCAACAGATCGAAGGTGGGTATGTTTAAATGGGAACAGATCCCTAG 1035
Db 5989 TGATGGATCAAAATACAGTGGGAGGTGGTATGTTT- - - - -AAGATGACAAATGATCTCG 6043
Qy 1036 AATTACTCCAATTTGGACATTTTCATACGCAAAAAACAAGTTTAGACGAGTTACCAAGTTTTT 1095
Db 6044 TGTTACAAAAATTCGTCGCTTATTCG- - - - -TAAACAGTTTATGATGAGTTACCCAGTTTT 6102
Qy 1096 ATAAATGTTTTAAATTTGGCGATAGTAGTCTAGTTGGTACAGTCCACCTACAGTTGATGAAT 1155
Db 6103 GGAATGTCTTTATAGGAGATATGAGTTTGTGGGAAACACGTCACCTACAGTAGACGAGT 6162
Qy 1156 TTGAAAAATATACCTCTGGTCAAAAGAGAGATTTGAGTTTAAACCGAGGATACAGGTC 1215
Db 6163 ATGATCAGTATCTCCAGAACAGAACGTCGACTCAGCTTTAAACCTGGTATTTACAGGTT 6222
Qy 1216 TCTGGCAGGTTAGTGGCTAGTAGTAATATCAAGACTTTCGACGAGTAGTTGGTTGGACT 1275
Db 6223 TATGGCAGGTTAGCGCGCTAGTAAATAACCGATTTGACGATGTTGTAATAATTAGATG 6282
Qy 1276 TAGCATATCATGATAAATTTGGATATCTGGTCAGATATTAATAATTTTAAAGACAGTGA 1335
Db 6283 TGGCTTATTTGATAAATTTGGAACAATCTGGAAGATATTTGAAATTTTGCCTTAAAACTGTTA 6342
Qy 1336 AAGTTGTATTTCTCAGAGAGGGAAGTAAGTAAGATATATGAAAGTTTGTGTCGGTT 1395
Db 6343 AAGTTGTATTTATGAGGAATGGAGCGGAAGTAAGG- - - - -ATGAAAGTATGCTAGTTGGAT 6399
Qy 1396 CTTCAGGCGGACATTTGACTCAGCTGTATTTGTTTAAACCGTTTGGGAAGGAAGAAGAAC 1455
Db 6400 CTAGTGGAGGTCATTTAACTCATTTGTTATTTTAAACCTTCTCGAAGATAGAGAAA 6459
Qy 1456 GTTTTTCGGTAACATTTGATAAAGAGGATGCAAGAGTCTTTTGAAGAAATGAATAATGT 1515
Db 6460 GAATTTGGGTGACTTTTGTATAAAGAAGATGCAAGAGTATTTTGAATAAATGAATAATTT 6519
Qy 1516 ATCCATGTACTTTTCCAAACAATCGCAATCTCATTAATTTAGTGAATAATACITTTCTTAG 1575

Db 6520 ATCCATGTTTATTCCTAATCGTAATCTAAAAATCTTATTAATAAATACGATTTCTTG 6579
Qy 1576 CTTTCAAAAATTTTACGTGATGAGAAACACAGATGTTATTTTTCATCTGGTGGCGCGTTG 1635
Db 6580 CTTTCAAAAATTTTGAAGAAAGAACGCCCTGATATTTATCGTCTCATCAGGGCGCAGCTGAG 6639
Qy 1636 CTGTCCTCTCTTTTATCATCGGAAACATTTTGGAGCAAGACGATTTATTTTGAAGTAT 1695
Db 6640 CAGTTCCTCTCTTTTATCTAGGGAATAATTTTGGTCTGAAGACAGTCTATATAGAAGTAT 6699
Qy 1696 TTGATCGAGTTAATAAATCTACATTAACCTGGAACCTAGTTTATCCGCTGACACATATTT 1755
Db 6700 TTGATAGAATTTGATCTCCGACAACTGACTGGGAAACTTGTGTTTATCCAGTACTGACAGGT 6759
Qy 1756 TTATGTTTCAGTCGGGAAGAAATGAAGAGGTATATCTTAAATCTATTAACCTTGGGGAGTA 1815
Db 6760 TTATGTTTACAATGGGAAGAAATGAATAAGTATATCCGAAGCAATCAATCTCGAGGGA 6819
Qy 1816 TTTTAAATGATTTTGTAAACAGTAGGACTCATGAACAACAGTTTAAATCGATGATAAA 1875
Db 6820 TTTTCTAAATGATTTTGTGACAGTAGGTTCTCATGAACAACAGTTTAAATCGACTTATAA 6879
Qy 1876 AGAGATTTGATTTATGAAAAAATAATGGAAGTATAACCGACGAAATATTTTATCAAAACAGG 1935
Db 6880 AGAGTTTGATAGATTAAGAGTGAGGATTTATTCAGGATGATGTTTATTCANACAGG 6939
Qy 1936 ATATCTGACTATATTCAGAAATATTCAGAGTATAAAAAATTTCTCAGTTTACAAAGAAAT 1995
Db 6940 ATACTCTAATTAATGTTCCAAAATTTTGTAAATCGGAATAATTAATCTTTATGAAAAAAT 6999
Qy 1996 GGAACAATATATAACAATCAGAAGTAGTTTATTTGCCAGGAGGCCCGCTACTTTTAT 2055
Db 7000 GAATCAATTTGATTAAGGAATCAGATATTTATCATTCCTGCGGTCAGCTACGTTTAT 7059
Qy 2056 GAATTCATATCCAAAGGAAAAACAATTTATTTGTTTCTTAGCAAAAAAGTAGTGTA 2115
Db 7060 GGAGTTTATGCTTAAAGTAAAAATCCAATTTGTTTCCGCGGTAATAAAAAATTTGTTGA 7119
Qy 2116 ACATGTAATGATCATCAAGTAGAGTTTGTGAAGAATTTTACAAGATAATAATATTTT 2175
Db 7120 GCATGTAATGATCACCAGATGCAATTTTGTAAAAATAACGAAGAATAATACAATTTAAT 7179
Qy 2176 ATTTATAGAAAAATATAGATGATTTTGTGAAAAATTTATTTGAAGTT- - - - -TCTAAGCAAAC 2232
Db 7180 AGTTATAGATGATATTTTCAGACTTACATTTAATTTCTCATAAATTTTAAAGGCAACAATTT 7239
Qy 2233 TAACTTTACATCAATAATAATTTTTCGTGAAAGATTTAAA- - - - -ACAAATAGTTGAAAA 2289
Db 7240 TGAACCTTATTCGAATAACGAGAGATTTAATGACGTTTCAATGTGGAATTCAGTAACCT 7299
Qy 2290 ATTTAATGAGGATCAAGAAAAATGAATAAAAAAGATGCATATTTGATATGCTTATC 2349
Db 7300 TTTTAAAGGAAAAATAAATAAATGAAATAGAAATAGAACTCAGTATTTTCTTTTATAAAT 7359
Qy 2350 ATAAATTTCTCAGATTTTACT 2371
Db 7360 ACCTTTGGTTTATTTATCTTACT 7381

RESULT 7

AY376403

LOCUS

DEFINITION Streptococcus agalactiae strain 7271 Cps7C, Cps7D, Cps7E, Cps7F, Cps7G, Cps7H, Cps7M, Cps7I, Cps7J, Cps7K, Cps7L, NeuB, NeuC, NeuD, and NeuA genes, complete cds.

ACCESSION AY376403

VERSION AY376403.1

KEYWORDS GI:38640633

SOURCE Streptococcus agalactiae

ORGANISM Streptococcus agalactiae

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.

| | | | |
|-----------|--------------------------------------------------------------------------------------------------------------------|--|--|
| REFERENCE | 1 (bases 1 to 14202) | | |
| AUTHORS | Cieslewicz,M.J., Glusman,G., Chaffin,D., Kasper,D., Madan,A., Rodriguez,S., Fahy,J., Wessels,M.R. and Rubens,C.E. | | |
| TITLE | Evolution of group B Streptococcus capsular polysaccharides | | |
| REFERENCE | Unpublished | | |
| AUTHORS | 2 (bases 1 to 14202) | | |
| TITLE | Cieslewicz,M.J., Madan,A., Wessels,M.R., Kasper,D. and Glusman,G. | | |
| JOURNAL | Direct Submission | | |
| TITLE | Submitted (28-AUG-2003) Channing Laboratory, Brigham and Women's Hospital, 181 Longwood Ave, Boston, MA 02115, USA | | |
| FEATURES | Location/Qualifiers | | |
| source | 1..14202 | | |
| | /organism="Streptococcus agalactiae" | | |
| | /mol_type="genomic DNA" | | |
| | /strain="7271" | | |
| | /serotype="VII" | | |
| | /db_xref="taxon:1311" | | |
| | 128..820 | | |
| | /note="capsular polysaccharide chain length regulator" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps7C" | | |
| | /protein_id="AAR25945.1" | | |
| | /db_xref="GI:38640634" | | |
| | /translation="MNKIANTVEVINPILNKLKWKKLTITVAIAFATAGLYSIF | | |
| | IVTPQYTSSTRYIYNPNTNPNSSITADQLQAGSLANDYKEIITSDVLEKVISSEKL | | |
| | NYPSSQLQKITVSLIKDTRVISISVEDANPKMSQKLANSVREAAVKIKAVTQVEDI | | |
| | TTLKGNLPKAPSSPNKKNVLIGFIVGAGLSTVLVIMGILDRVNTTEEDIEKVLGI | | |
| | TSLGIVLPDLNKL" | | |
| | 831..1529 | | |
| | /note="capsular polysaccharide chain length regulator" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps7D" | | |
| | /protein_id="AAR25946.1" | | |
| | /db_xref="GI:38640635" | | |
| | /translation="MTLREIVDSKLRQAKTTEFVNAIRTNIOFSKENKILATISVR | | |
| | EGEGKSTASTSLAQLAGFKTLIIDADTNSVMSGTFFKATGIGITNYLSGNADL | | |
| | GDICETNPLMLVPPSGKVPNPNTALLQNAFNKMEIAIKNIIDYIIDPTPLGLV | | |
| | DAALIANACDGFILVTQAGRIKRYNVEKAKQMSQSGKFLGIILNKVNESVATYGDY | | |
| | GDYGNKQRDRKK" | | |
| | 1581..2930 | | |
| | /note="putative glycosyltransferase" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps7E" | | |
| | /protein_id="AAR25947.1" | | |
| | /db_xref="GI:38640636" | | |
| | /translation="MIQTVVVYFSASLTLTITPNFKNKDLFLVLLIHYIVFVLSDF | | |
| | YRDFWSRGYLEEFKRWLVKYSFYIPISSSLFFIPKNSFTWRLSFFFPFIAMNSILLYL | | |
| | LNPLKYVRKYSYAKFRSDTKVLIITNKDSLKMTFKNKYDHNYIANVILDSKDCY | | |
| | DLKNSLRILNKDALTSBELTCLTVDQAFINIPILFGYQIQDIINDIEANGVIYVNV | | |
| | VEALSFNIGBKRIQTGEYSVVTYISMKFYKSHLIARFIDIMGAILGLTIGIVAI | | |
| | FLVPOIRNKGPAIFSQNRVNGRIIRFYFRSMRVDAEQIKDLVHNMOTGMIFK | | |
| | LDDPRTIKIGIKTRKTSIDELPOFYNVLKGMISLVGTRPTVDYEYKNSYTKRRLS | | |
| | FKPGITGLWISGRNNITDFEIVKLDVQYINENSISWSDIKIILLTKVLLGTGAK" | | |
| | 2954..3403 | | |
| | /note="beta-1,4-galactosyltransferase enhancer" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps7F" | | |
| | /protein_id="AAR25948.1" | | |
| | /db_xref="GI:38640637" | | |
| | /translation="MKICLVSSGGHLAHLNPLKSIWEKEDRFWTFDKEDARSILRE | | |
| | EIVYHCFPTNRNVLNKLITLAPKVLKRPDPVLISSGAAVAPFPFYIKLFGCKT | | |
| | IYIVFVRIDQITPTGKLVVPTVDKFIQWEMKKVPAKINLOGIF" | | |
| | 3403..3894 | | |
| | /note="beta-1,4-galactosyltransferase" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps7G" | | |
| | /protein_id="AAR25949.1" | | |
| | /db_xref="GI:38640638" | | |
| | /translation="MIFVTVTGTHQFNRLRIKEVDRLKGKTGAIDQEVFIQTGYSDPEF | | |
| | QNCWSEKFLSYDDMNCYREAEIVITHGGPATFNNVAVSGKTKTIVVPRQEQFGHVVN | | |
| | HQVDFNVKVKLIYNFDIIDIETGQLNVYEGAMNRPLETNNSNFIKPKKILKELCD | | |
| | ENO" | | |
| | 3881..4975 | | |
| | /note="capsular polysaccharide repeat unit polymerase" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps7H" | | |
| | /protein_id="AAR25950.1" | | |
| | /db_xref="GI:38640639" | | |
| | /translation="MKINKSLFYIAFLVNFPSKLSGLGSENSYKIYVLAAILCAL | | |
| | KFLDLSYFERRKLVITLFLFVATILNLFVHKVTFFILTLIFSALKADISLKAPASII | | |
| | IGSRILGVLNQLIYFVLDVFIKYVNFYRDGHFILRSLDGLGHNFHNFALTFPLY | | |
| | IVLNYKRLKLIIVLITLNYLQYTFSTGYIVILFVILYLTNLSLKRVRFMKL | | |
| | APYIQFLAFTPLSSTIFPNSFVQKLDKLLTGRLNVAHLQLDGLTLFGNSFKETS | | |
| | VLFDNSYSMLLSMYGVLTLCMLIIVYVSKIRIVELQQLLLFIMSIVLFTSFYPSI | | |
| | VMAISWLVFGKIFCDGIEPIKKEFTIIDI" | | |
| | 5013..5738 | | |
| | /note="beta-1,4-glucosyltransferase" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps7M" | | |
| | /protein_id="AAR25951.1" | | |
| | /db_xref="GI:38640640" | | |
| | /translation="MIPKVIHYCWFGGNPLDNLKKYIKTWREQCDDYIEIEMWERNY | | |
| | DVTNKIMRDAYTKKNFAYSDVARLDIIVYGGFYLDTOVELLKSLLDPLRVHNCFLA | | |
| | RELSYDNTGLIIGAIRGOKFIDKNMSVREESKFLSPDKTCVEITTSLLAKKGLKSN | | |
| | LIQVIDGYIIPROYFNPKNLLTGKLDLINDTYSIHHYEGSWKKPSIFPDSMKIRVR | | |
| | LVIDFYGTYRRLRLRLFLKLQ" | | |
| | 5782..6744 | | |
| | /note="beta-1,6-N-glucosaminyltransferase" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps7I" | | |
| | /protein_id="AAR25952.1" | | |
| | /db_xref="GI:38640641" | | |
| | /translation="MKVLSLIPVYNGAPIETLESWATTYSNLGNINCCDSDTYT | | |
| | SVICQKTSKDKRKYKENGOGSARNLGVLAEGDWIAFLDSDDLCSFPYIBLV | | |
| | RVQSETGADIVITGIEDRPNHMEKLTCDKSTYSTLSVEATLTMYDNYNGISPLGKL | | |
| | YSKALLRHPYPIGKIHEDDTDTYKLVASSTRIAPSKETVYIKQPNSTTQSYREN | | |
| | MLYFDAMQHRDPFILRSKDELIDALNYTIVGGMWIVNKLLEVDKYDYISUREN | | |
| | YRKYKSLFNKRVITIKKIKYLSFISMMNLFLEFKIYNYSEG" | | |
| | 6748..7713 | | |
| | /note="beta-1,4-galactosyltransferase" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps7J" | | |
| | /protein_id="AAR25953.1" | | |
| | /db_xref="GI:38640642" | | |
| | /translation="MKKISVIIPVNVQSFINECIDSVLAQTSNLEIILVDDGSDTD | | |
| | SGDICDYKDKRILVPHKNGSLDARKKENSRAATGDIYLLSDSDLYLKEDAIEE | | |
| | VAHGKYSRSDIVLACIVERGRIINIVILENETIETISPLQAIRNYIYDODYKRAIF | | |
| | MVNNKYLRELFSTMCYPCVQKHEDEFLTYKLAKNRIIPFRYTYDNYIYRESIMT | | |
| | SVNKKELHAIIEALQERHILKEYPFVQSEKALIKTMEVNLTEKNDYKPYEFTLL | | |
| | RNEYORTLHIEKQGVLLKYSLCYILYFKILNLYKRRKQNMNK" | | |
| | 7710..8666 | | |
| | /note="alpha-2,3-N-acetylneuraminic acid synthetase" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |
| | /product="Cps7K" | | |
| | /protein_id="AAR25954.1" | | |
| | /db_xref="GI:38640643" | | |
| | /translation="MNYFVCHTLYHLITIKDKENTRIFICDTIPDYCAWIKILN | | |
| | NQGIKTLFELDELDYRQKNNEEVMNLVDSLYKHFERVNTQVILFNDITLMGRYM | | |
| | AYLGKNTLIEDJEDNCFQAKLFFGVSVKRIIKTYFLKKYVYGFYSKYCUSIENSL | | |
| | GLPHDRSKYKELRFLDLSQEQSLIFKLFKMKPTIAPKSVLLTQPLAQDK | | |
| | WYKTATERFOSIQOYDYFEGIVHDYRERGNYILKVHPRDAVYSKLPVELLPSNIP | | |
| | MEIIEMLTGTGFEVGIHSTALDFLTCVDRKIILVNLKDIK" | | |
| | 8663..10063 | | |
| | /note="capsular polysaccharide repeat unit transporter" | | |
| | /codon_start=1 | | |
| | /transl_table=11 | | |


```
/product="Cps7L"
/protein_id="AAR25955.1"
/db_xref="GI:3864064"
/translation="MNTKLLQSGFIYITIGNLLVOGLAFITLPIYTRVISTEVYQVS
LYVAMGILFIQGLQSGSSARVYGBEFKSYGSFSGVNGINFLVLLVAFUPR
SFLAPVDFSEISFLIMICOSYANYVYVFFQYPIQQOQSFANLILSLTNVASSVALS
LFLIFHWSDDPFLSFGAFVPTIIVAIYAFYIYHKSFPNGSINMAMIPWFFDARK
HLGHQILNOLDRLMLARLYNTKEVAMVSPGSLGMIIQIVLNSINMAMIPWFFDARK
AKLQLPYISRYLYGLVFLTLGLYLTIPBELAQIMGDKYSSSVQFSLIISYFLVF
LYPFPVNIQFYANTWIPGTLAAGVNMLLNLVLPNTAAYGAAMWATVSYLALLI
FHHIVSKYKHYLDVSRQXIIISGIVFSYAMLMNMFLDNIVVRWGLGIMILIVSIV
FOKVLDLLSKRRRR"
1063. .11088
/notes="N-acetylneuraminic acid synthetase"
/codon_start=1
/transl_table=11
/product="NeuB"
/protein_id="AAR25956.1"
/db_xref="GI:3864064"
/translation="MVVYIAEIGNHNGDINLAKMVDVAVSCGVDAVKFOTFKAELK
ISKPAKAEYQKATGTGADSOLEMTKELESFEYLEIRDYALSKGVETESTPDRBS
LEFLISLDMPIYKIPSGEINLNPYLEKIKQKQKVIILSTGMVNEEIHQVAINRQNG
TDSILHCTTEYPTPYPSNLNVIHTLKDFKDLITIGYSDHSISGVPYIAAAGAB
VIEKHFTLDNMEGPDHKASATPDILALVKGVRIVQALGRFEKIPDPVEERKNIYA
RKSVALKPRIKKGIYSIENITVKRPGNGISPMNWDILGQEAQDDPEDEVIRDSR
ENQLPEL"
1165. .12319
/notes="UDP-N-acetylglucosamine-2-epimerase"
/codon_start=1
/transl_table=11
/product="NeuC"
/protein_id="AAR25957.1"
/db_xref="GI:3864064"
/translation="MKKICLVTSRABYGMKPLIQLSKDKENVLQIATAMHLEBK
YGYTRQIEDGDIAYKVLPHLYDTRVSTAMHLQGLTKIFDKEDYDLVILIG
DERMLPVNVALIYVPVCHLGGGETSLGNDFEYIRHAIITKMSHLHLVSTEDPRQV
IOMGEOPQVINTGALGVENALSIPSLTKEAIEKQLGVLSESVFVLYHPVTPBOQK
SAGQMKVIALSLSKFGVQCLFIGNSDSTGSDDIKAINTLYLINHNSYCFASISLQ
YSLIRHSLGLIGNSSGLLEVPSLMKPTLINIGRQKRLHGESVSVVPVETSVLEG
LSKUNEVNFDNPIYGNASSIAYEAIKLYLKDEPISQPPFYDLKENNLK"
12316. .12945
/notes="putative acetyltransferase"
/codon_start=1

Query Match 12.0%; Score 839.6; DB 1; Length 14202;
Best Local Similarity 48.6%; Pred. No. 7.3e-86;
Matches 2850; Conservative 0; Mismatches 2949; Indels 71; Gaps 17;

Qy 16 TGGCAATATTGATGATAGCAGTTGCAATTTCTGCAATCTTAAACAGTCATATACCA 75
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATATATAATGATGGTTCATPATTTGCAT 132
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1636 CTCCCAATTTTAAAGCAATAAAGATTATTGTTTGTCTTCTATGATACATTATATTGCT 1695
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 133 TTTTATATCTCGTATGCCAGTTGAAATTTAGTATAGAGGTAATCTGATAGAGTTTGAAA 192
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1696 TTTATCTTCTGATTTTACAGAGACTTTTGGAGCTCGTGGCTATCTTTGAAGAGTTTAAA 1755
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 193 AAACATTAACTATAGTATAATATTGCAATTTTCTTACGGCAGTATCATTTTGTGTTGG 252
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1756 TGGTATTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCGTTATTTTATTTTA 1815
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 253 AGAATAATTTCCGACTTTCAAGACGTTGGTGGCGGTGTAFTTTCACATAATTAACCTCGTT 312
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1816 AAAACTCATTTACAATGACACGACTTTCCTTTTTCCTTTTATTGCTATGAATTCGATTT 1875
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 313 TGGTATACCTAATTAACGTAATTAATAGCAGTTTAAGGATAGCTTTCTATTTCGACAA 372
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1876 TATTATATCTATGAATTCATTTTTTAAATAATATTATCGAAAAATATCTTACGCTAAGTTT 1935
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 373 TCTATCAAAAAAAGACGATTTCTAATTAACAAGGCTGAACGATGGGAAAAATATGCAAGTTT 432
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 1936 CACGAGATACCAAAGTTGTTTTGATAACGAATAAGGATCTTTATCAAAAATGACCTTTA 1995
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 433 TATTTGAATCACATAAACAATAATCTTTGTCATTTGGTAGTTTGTAGGTACAG 492
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1996 AGAATAAAT---ACGACCATAAATATATCGTCTCTGTATATCTTTGGACTCTCTGAAAAGG 2052
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 493 AAATAGATAAAATTAATTTATCATACCCTCTATTATTCTGTGGGAAGAGCTATAGAGT 552
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2053 ATTGTTATGATTTGAAACAATACTCGTTAAGGATATAAACAAGATGCTCTTACTTCAG 2112
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 553 TTTTCAACAAGGGAAGTGGTCGACCACGCTCTTTATAAATCTACCAAGTGAGTTTTTAGACG 612
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2113 AGTTAACTCTTAACCTGTTGATCAAGCTTTTATTAACATACCATTGAATTAATTTGGTA 2172
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 613 TAAAGCAATC-----GTTTCAGATTTTCAGTTGTTAGGTATTTGATGTAAGCGTTG 663
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2173 AATACCAATACAAGATATATTATTAATGACATTTGAAGCAATGGGAGTGAATTTGTCAATGTA 2232
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 664 ATATTAAATTCATTCGGTTTTTACTCGGTTGAAAAACAAAAATCCAACTGCTAGGTGACC 723
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2233 ATGTAGAGGCACTTAGCTTTGATAATATAGGAGAAAGCGAATCCAAACTTTTGAAGAT 2292
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 724 ATAGCATTTGTAACCTTTTCCACAAAATTTTATAAGCCTAGTCATATCATGATGAAACGAC 783
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2293 ATAGTGTATTACATATCTATGAAATCTTATAAATATAGTCACCTTATAGCAAAACGAT 2352
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 784 TTTTGAATATACCTCGAGCGGTAGTCGGGTGTAATTAATTTGTGTTAGTTTCTATTTTGT 843
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2353 TTTTGGATATCATGGGTGCTATTATAGTTTGTGTCATATGTGGCAATTTTGTGCAATTTTTC 2412
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 844 TAGTTCCAATTTATTCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAAACGAGTTGAC 903
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2413 TAGTTCGCAAAATCAGAAAAGATGGTGGACCGGCTATCTTTCTCAAAATAGAGTAGGTC 2472
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 904 AGAATGGACGCATATTTTACATTTCTCAAGTTTCGATCGATGTATGTTGCTGCTGAGGAGC 963
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2473 GTAATGGTAGATTTTGTAGATTTCTATAAATTCAGATCAATCGCGAGTAGTCGAAACAA 2532
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 964 GCAAAAAGACTTGTCTCAGCAAAAACAGATGCAAGGTGGGTATGTTTTTAAATGGGAA 1023
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2533 TTAAGAAAGATTTATTAGTTTCACAATCAAAATGACAGGG-----CTAATGTTTAAAGTTAGA 2587
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1024 AAACGATCTCAGAAATTTACTCCAATTCGACATTTTATACGCAAAAACAAAGTTTACAGAGT 1083
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2588 CGATGATCCTAGAAATTTACTAAAAATAGGAAATTTATTTCG-AAAAACAAAGCATAGTAGT 2646
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1084 TACCACAGTTTATTAATGTTTTTAAATTTGGCGATATGAGTCTAGTTGGTACACGTCACCTA 1143
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2647 TGCCTCAATTTCTATAATGTTTTTAAAGGTGATATGAGTTTAGTAAACACGCGCTCCCA 2706
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1144 CAGTTGATGAAATTTGAAAAATATCTCTGGTCAAAAGACAGATTCAGTTGAGTTTAAACAG 1203
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2707 CAGTTGATGAAATGAAAAAGTATAAATTCACGCAAGAGCGACGCTTTAGTTTAAAGCAG 2766
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1204 GGAATACAGCTCTCGCAGGTTAGTGGTCTGTTAGTATATACAGAGCTTCGACGACGTAG 1263
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2767 GAATCCTGTTTTGTGGCAAAATCTGTTAGAAATAATATTTCTGATTTTGTGAAATCG 2826
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1264 TTCGTTGGACTTTAGCATACATTTGATAATTTGGACTATCTGGTCAGATATTTAAATTTTAT 1323
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2827 TAAAGTTAGATGTTCAATATATCAATGCAATGGTCTATTTCGTTCAGATATTAAGATTATTC 2886
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1324 TAAAGACAGTGAAGTTGTTATTTGTTAGAGAGGGAAGTAAGTAAAGTA----- 1372
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2887 TCCTAAACATTTAAAGGTAGTCTTACTTTGGGACAGGAGCTAAGTAAAGGTTTGAAG 2946
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1373 -----TATCAAGTTTGTTCGTTCCGTTCTTCAGGGGACATTTGACTCACTGTATTT 1426
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2947 GAATATAATGAAATTTGTCTGGTTGGTTCAAGTGGTGGTCTATCTAGCACACTTGAACCT 3006
Qy ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1427 GTTAAAAACCGTTTTTGGGAAGGAAGAAACGTTTTTGGGTAAACATTTTCATAAAGAGGATGC 1486
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


Db 3007 TTTGAAATCCATTTTGGGAAAAGAGATAGGTTTTGGGTAACTTTTGATAAAGAGATGC 3066
Qy 1487 AAGAAGTCTTTTGAAGAATGAAAAAATGTATCCATGTTACTTTTCAACAAATCGCAATCT 1546
Db 3067 TAGGAGTATCTAAGAGAAGAGATTGTATATCAATTCCTTTTCCACACACCGTAATGT 3126
Qy 1547 CATTAATTTAGTGAATAATACCTTTCTTAGCTTCAAAATTTTACGTGATGAGAAACCCAGA 1606
Db 3127 CAAAACCTTGGTAAAAAATCTATTTCTAGCTTTTAAGGTCCTTAGAAAAGAAAGACCAGA 3186
Qy 1607 TGTATTATTTTCACTGCTGGCGGCTTGTGTCCTTTCTTTTACATCGGAAACACTATT 1666
Db 3187 TGTATCATATCACTGCTGGCGCTGTAGCAGTACCAATCTTTTATTTGTTAGTTAAT 3246
Qy 1667 TGGAGCAAGACGAATTTATTTAGAGTATTTGATCGAGTTAATAAATCTACATTAACCTGG 1726
Db 3247 TGGTTGTAGACCAATTTATATAGAGTTTTCGACAGGATAGATAAACCACTTTGACAGG 3306
Qy 1727 AAAACTAGTTATCCCGTAAACAGATATTTTATTTGTTGAGTGGGAAAGAAATGAAGAAGT 1786
Db 3307 AAAATTAGTATCTCTGTAAACAGATAAATTTATTTGTTGAGTGGGAAAGAAATGAAGAAGT 3366
Qy 1787 ATATCTTAATCTATTAATCTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAT 1846
Db 3367 TTATCTTAAGGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTACACAGTGGGACA 3426
Qy 1847 CATGAACCAACAGTTTAAATCGATTGATAAAGAGATTTGATTTATGAAAAAATAATGGAAGT 1906
Db 3427 CATGAACAGAGTTTCAACCGCTTATTAAGAGTTGTATTAAGAGGACAGGTGCT 3486
Qy 1907 ATAAACGACGAAATATTTATTAACAAACAGATATCTGACTATTTTCCAGATATTTGCAAG 1966
Db 3487 ATTGATCAAGAAGTGTTCATTTCAACCGGTTACTCAGACTTTTGAACCTTCAGAAATGTGAG 3546
Qy 1967 TATAAATAATTTCTCAGTTACAAAGAAATGGAACAATATATTAACAATCAGAGTAGTT 2026
Db 3547 TGGTCAAAATTTCTCTATATGATATGATGAATCTTTACATGAGAGAGCTGAAATCGTT 3606
Qy 2027 ATTTGCCAGGAGGCGGCTACTTTTATGAATTCATTTATCMAAGGAAAAAACAATTA 2086
Db 3607 ATTACATATGCGGCTCAGCGAGCTTTTATGAATGACGTTTCTAAGGGGAAAAAACCAATA 3666
Qy 2087 TTGTTTCTAGACAAAAAAGTATGTTGAACATATGATAATGATCATCAAGTAGAGTTGTA 2146
Db 3667 GTAGTTCCAGACAAAGACAGTTTTCGAGAGCATGTGAATATCATCAGTAGATTTTGT 3726
Qy 2147 AGAAGATTTTACAAGATAAATAATTTATTTATAGAAAAATATAGATGATTTGTTGAA 2206
Db 3727 AATAAGGTCAAGTTGATCTATAATTTTGATATTTATATAGACATTTGAAGGGTTACAAAAT 3786
Qy 2207 AAAATTTATGAAGTTTCTAAGCAAACTAACTTTTACATCAAAATAAATTTTGT 2263
Db 3787 CTAGTTTAAAGGAGCGATGATCGTCGTTTTTATAGAAACCAACATAGTAATTTAT 3846
Qy 2264 GAAAGATTTAAACAAATAGTTGAAAAATTTTATGAGGATCAAGAAATGCAATAATAAATA 2323
Db 3847 AAAGAAATTTAAGACATATTAAGAGGTTGTGTGATGAATCAATAAATACTTTAT 3906
Qy 2324 AGATGCAATTTGATAATGCTTATCAATAATTTTCTCAGATTTTACGAGAGGATAC 2383
Db 3907 TTATATTGCAATATTTTAGTTAAATTTTAAATCACTAGGTTTA - GGAGAGGAAAAAT 3964
Qy 2384 AGATATTATCATCTTCTCAGGAGAAATGCAACCAATAGTTCTTCAGATACCTGTAT 2443
Db 3965 TCAACTTACAAATAGTAATGTAGTTGCAATCTTTATTTGTGCAATAAATAATTTTACTC 4024
Qy 2444 AATTATTTTAAATATTTCTCAGGATTTATATGTTGAATTTTACAAAGATGAGCAAAAATAT 2503
Db 4025 GATAGCTTTATTTTGAAGAAGAAAACTCGTTACTATCTCTCTGTTATTTGTCGGACT 4084
Qy 2504 AAAAGAAATPAGGATATATGAACGAGTTAAATGTTTACAGATTTTCTTAATATATCAGAA 2563
Db 4085 ATTTTGAATTTATTTTGTACATAAGGTTACTTTTATATTAATTAATTTTCCCTA 4144

Qy 2564 AAAACTATGTATAATGTACTGTTTAGAATTTTATTAAGAATGTATCGAGCTTTTGAATAC 2623
Db 4145 GCAUTTGAAGGATATCTCTCTAAAAAAGCTTCTCTATAATAATAGGATCGGTATATG 4204
Qy 2624 TATTACAAAGATTTGTTGTTTATTGATAGATAAATAACATGGTCTAAGAATAAGATTG 2683
Db 4205 GGAGTTCTATTAAATCAAAATTTTGTGAATTTAGATTTTGTAGAATAAAGTATGTCAT 4264
Qy 2684 GTTCTAATTTGGGTTTCGCTTCCACATGATTTTGTGGCAATTTCTTTTATCAAAATGAAAAAG 2743
Db 4265 TTTTATAGAGATGACATTTTCATCTGAGAAGTGAATTTAGGTTTGTGGCATCCCACTTT 4324
Qy 2744 AAAACAGCTTATTATTTAAAGTAATCTAAATGTCGAGATGAACATAATTATACAGACAAATTA 2803
Db 4325 ATTCATAATTTTGTCTCTAAACCTTTTGTGATATATGTTACTCAATATATAAAGCACTG 4384
Qy 2804 TAGAAAAATAGAAATTTTCAAAATAGATATCTAAATATGGAATTTAAGA - - - - -TATA 2857
Db 4385 AAGCTTATTGTAATGTTTATTTTAACTTAATTAATTTTACTGATCAATATATCTTTT 4444
Qy 2858 TAAAGTGGAAAAAATCAACATCTTCTCTATGCTTTTACAGATGATTTCTATTGATGAAT 2917
Db 4445 TCAAGAACAGGGTATTTATTTGTAATTTTATTTATTTGTTACTCATTTATCTAACAAAGAT 4504
Qy 2918 TGTAAATGCAAGAAATTTAGGTTTTTATTTGCTAGAAAGTTAAAAATGAAAAATAAAT 2977
Db 4505 AGTCTAATAAAGAGATTTTATGAAATTAGCACCATAATACAGTTTTTTTATTTATAGCA 4564
Qy 2978 CTAAATTTAAAGAAATTTATTTACTAAAAAATAAATAGTTTGTGAGAGTATGAT 3037
Db 4565 TTTACTTTTTGAGTCTCAAAATTTTTTCAATTTTGTTCAAAAATTTAGATAG 4624
Qy 3038 GTTTAAATTTTAAATATGACCCGGAATTTTATTTTAAAGTACTTTCTGTTGATTTAT 3097
Db 4625 CTTTTGACAGTAGTTAACTATGCTCATTTACACTTTGAGCGGTTGACTCTTTT 4684
Qy 3098 TTTTATTTCCAGACAAAGTATGTTATTTTATTTTATTTTATGAAATTTTAAATTTTCA 3157
Db 4685 GGAATAGTTTTTAAAGGAAACCAAGTCTGTTTGATAATAGTACTCTATGTTA - TTGA 4742
Qy 3158 TATTAATTTTGAATAAATAAGCTAAATTAATAAATAAATAAATTTTATTTGTTTATTTATG 3217
Db 4743 GTATGATGTTGTTAGTCTCTTACCTGTTTGTATGATTAATCTATTATGTTTACAGTAAG 4802
Qy 3218 GTCTATATTTATGTTTGTCTCAGTAGTCACAGATGTTGTTGGAATAAATTTTGAAG 3277
Db 4803 AAATAAGAAATAGTTGAACCTTCAATTTACTGCTATTTATTAATGTCTATAGTATTTTACAG 4862
Qy 3278 ATTTATTTGCAGATTTTACTGCTCCCATAAATTTGGATTTATGCAATAATGATTTATAATTT 3337
Db 4863 AGAGTTTTTACCAGATAGTTATGATATTTAGTTGTTAGTTTTCGGAATAATTTT 4922
Qy 3338 GTATTCATTTATAAATATGATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3397
Db 4923 GTGATGTTTGAACCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4982
Qy 3398 AGTTTATTTAGGTATATCTGCTATGTTATTTTCAAAATCGGAAGATATTTGATTTT 3457
Db 4983 ATTCGATATAGTAGGATGTCAGGAAACAAATGATACCTAAAGTTTATACATTTATTTGTTGG 5042
Qy 3458 AGACAGACACCTTATAGGACTAGACTATCTTTATTAACAGCGCTCAAAACAAAGGTTGTTGG 3517
Db 5043 TTTGGAGGAATCCCTTACCTGATTAATTTAAATAAGTACATAAATAAATAAATAAATAAATA 5102
Qy 3518 CTTTATGAATCTCTACGTTTAAATACCACTACAAATATAGTTTCAATTCGGTTAATCTT 3577
Db 5103 TGTCCAGATTACGAAATTTATTTGAATGGAAACGAAATTAATGATGTTTACCAAAAAATTT 5162
Qy 3578 TGCATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3637
Db 5163 TTTATGAGAGAGCATATATCTAAAGAAAAATTTTGCCTATGTTTCTGATTTATGCAAGGTTG 5222

| | | | |
|----|------|-------------------------------------------------------------------|------|
| Qy | 3638 | CTATTTAAGTGGATCGAGAAATGTGGTAGTATTATCGCTAGCAATATTAATATATATGCTTTGTT | 3697 |
| Db | 5223 | GATATTTATACACTTATGGCGGATTCCTCTAGACACTGATGTGGAACCTTTTAAAAAGT | 5282 |
| Qy | 3698 | ATGAGATATATAGGTGGAAAAATTTGCTTGGATAAAAAGCTAATAGTATATATTTGTAAT | 3757 |
| Db | 5283 | TTAGCCCTTTGAGAGTTCAACTGTGTTTTTAGCTAGGAGTTAAGTTATGATGTGTAAT | 5342 |
| Qy | 3758 | AC--TACTATTATTTTAAAACTCTGAATTCGCTTTACCATGAAATTTTGGCTGTTTATAAT | 3815 |
| Db | 5343 | ACAGGACTAATATTGTGTGCATTTAGAGGACAAAATTTATTAAGACAATATGTCGGTA | 5402 |
| Qy | 3816 | TCTAGAGAATCAAGTAAACGAAGCTAGATTTATTATTTTATCAAGGAAGTATTGATAAGTA | 3875 |
| Db | 5403 | TACGAAGAACTCAAAATTTTAAAGTTTGTATAAAAACATGTGTAGAAATCACGACCAGCTTA | 5462 |
| Qy | 3876 | TTAGAAAAACAATATTTTATTTTGGATATGGAAATATCCGAATATTTCAGTTACGGGAACCTGG | 3935 |
| Db | 5463 | CTTTTAAAAAAGGACTGAAAGTAAAAATATCATACAAGTTATTTGATGGTGTCACTATT | 5522 |
| Qy | 3936 | CTCGGAAGTCAATTCAGGCTATATATCATTTTTTATAAATCAGGAATAGTTGGTGTGATT | 3995 |
| Db | 5523 | TATCCAAAGACATACTTTAATTCCTAAAAATTTGTGACAGGAAGCTA-----GATTTTC | 5577 |
| Qy | 3996 | TTACTGATGTTCTCTTTTTTTTATGTTATAAAAAAAGTTATGAGGTAAATGGGAAACA | 4055 |
| Db | 5578 | TACAAATGATACTTATTCTATATACATCATTTATGAAGGAAGTTGGAAAAAGCCATCTATT | 5637 |
| Qy | 4056 | GCATATTTTATTTTACATCATTTAGCCATATTTTTCATATATATGMAACAATAGATCCGATT | 4115 |
| Db | 5638 | TTCTCGATTTCGATGAAAAATTA-GAGTAAAGACTTGTAAATGATTTTATTTTGGATATGGA | 5696 |
| Qy | 4116 | ATTATTATATTAGTACTATTCTTTCTTCAATAGGTATTTGGAAATATAATAATTTTAAA | 4175 |
| Db | 5697 | ACATATAGAAGCTGTTCGGTTTTTTTAAAGTTTAAACAGTAGTCGCTATTAACTGAAATA | 5756 |
| Qy | 4176 | AAGGATATGGAGACAAAAAATGAATGATTTAAATTTTCAGTTATTGTACCAATTTATATATGT | 4235 |
| Db | 5757 | TTTTCTTTAAAGAAATGGAGGAACAATGCAAGTATCATTAGTCATACCAGTTTATATATGG | 5816 |
| Qy | 4236 | CCAAGATTATCTGATATAATGTTATAACAGTATTTTAAACCAACAATATATCTAATTTAGA | 4295 |
| Db | 5817 | GGCACCTTACATCGAAGAAACATTTAGAAAGTGTCTATGGCCACAACTTTATTCAAATCTAGG | 5876 |
| Qy | 4296 | GGTTATTTCTCGTAAATGATGGAAGTACTGATGATCTCTGAGAAAATTTGCTTTAACTATAT | 4355 |
| Db | 5877 | AAATATTAAATTTGTTGATGCTCAACTGATTAACATCATGTTATATGCCCCAAATAACAC | 5936 |
| Qy | 4356 | GAAGAACGATGGAAAGATTAAATATTACAGAAAATTTAAATGGCGGCTCTAGCAGATGCTCG | 4415 |
| Db | 5937 | TTCTTAAGATTAACCGGTGAATTTTATATAAGAAATTTGGCGGTCAATCTCTGCTAG | 5996 |
| Qy | 4416 | AAATTTTCGGACTAGAACATGCAACAGGTAAATATTTTGTCTTTTGTGCGATTCTGATGACTA | 4475 |
| Db | 5997 | AAATTTAGGAGTCAAACTAGCTGAAGGAGATTGGATTGCTTTCTTTGGAATCCGATGACCT | 6056 |
| Qy | 4476 | TATAGAAGTTGCAATGTTTCGAGAAATGTCATGATAATATAACTCAGTATTAATGCCGATAT | 4535 |
| Db | 6057 | CTGCTCACCAATTCATATAGAAATTTATTAGTAAAGAGTCGAGAGTGAAACTGGTGTGATTT | 6116 |
| Qy | 4536 | AGCAGAGATAGATTTTGTGTTAGTAGACGAAAAACGGGTATACAAAGAAAAAAGAAATAG | 4595 |
| Db | 6117 | GGTGATAACTGGAATCGNAGACCCGCCACCATAGAATGGAAATACTTACATGTGATAA | 6176 |
| Qy | 4596 | TAAATTTTCATGCTTAAACGAGAGAAGACTGTGTAAGAAATTTTGTGAGGATCTAATAT | 4655 |
| Db | 6177 | GTCAACTTACTCAACACTCAGTGTGGGAAGCGACATTAACAGAAATGACTATATGATACTA | 6236 |
| Qy | 4656 | AGAAAAATAATGTTTTGGTGCAGCTTTTATTCAGGAGATTTATATAAAGATATAAATTTCCA | 4715 |
| Db | 6237 | CGGTATTTCTCCAATGGGAAAGTTATATCTTAAAGCAATTTATTTAAGACATCATCATACC | 6296 |
| Qy | 4716 | AATTAATAATAGAAAGTATTTGTTGAGGATTTGCTTTTTTAAATTTGAGGCTTTTGAAACAATGT | 4775 |

| | | | |
|----|------|------------------------------------------------------------------|------|
| Db | 6297 | TATTGGTAAAAATTCATGAAGATACAGATACAACTTACAAACCTGGTTGCAAGTAGTACACG | 6355 |
| Qy | 4776 | AACACGCTGTAGTAGTTGATACAGAGAAATATTATTATAATTATGTCTCATCGTAAACAGTTTC | 4835 |
| Db | 6357 | AATGGCGTTAGCAAAAGAGAAAACAGATATATATAAAAGGTCAACCTTAATAGTAGCAGCTAG | 6416 |
| Qy | 4836 | GCCTTATTAATCAGAAATTCCTATAAATAATATTTGATTTAGT-CACAAGATTTGGAGAAAT | 4894 |
| Db | 6417 | GCAGTCATATAGGAGAACATGTTATATTTATTTGATGCAATGCAACATAATAGAGACTT | 6476 |
| Qy | 4895 | ACCCCTTTAAGTTAAAAAGAGAGTTTGTGTCATTTATTTTGATGCAAAAGTTTATTAAGAGA | 4954 |
| Db | 6477 | TATTCCTTGAACGTTATTCAGACAAAGAACTAATTCGATGCTTTTAAATTTATAAAACTGTAA | 6536 |
| Qy | 4955 | AGGTTAAATGTTTAAACAAAATGCTATTTCAACAGATGTTTGGATTAATGAGTTCTTGCCAA | 5014 |
| Db | 6537 | TGAGAGATGTGATCGTTTAATTAAGCTTCTAGAAAGTTGATAAATAGCATATATTAATTC | 6596 |
| Qy | 5015 | TATTAGAGTCTTATCGAAAGAAATACGTAGATATCCATTTATTTAAAGCGAAAGATATT | 5074 |
| Db | 6597 | TTTGAGAGTGAATATAGAAATATTTACAGA-----GTATATTTTAAACCGTG | 6649 |
| Qy | 5075 | TATCAAGAAAGCAATTTAGTTTACGTGTAATTTGATGAAATTTTCGCCCTAAACTATATGTAA | 5134 |
| Db | 6650 | TGACAAATAAAGAAAAATAAATAATCTGTCAATTTTCTCAATGAACCTGTTTTTAT | 6709 |
| Qy | 5135 | TGTTATATAAGAAATTTCAAAGCAGTAGNGGTAA- AATGGATAAAATTTAGTGTATTTCG | 5193 |
| Db | 6710 | TCCTTAAGAAGATTTTCAACTATAGTGAGGGATTAATTTATGA AAAAGATCAGTGTATTATTA | 6769 |
| Qy | 5194 | TTCCAGTTTATATGTAGATAAATATTTAAGTAGTTGTATAGAAAGCATTTATTAATCAAA | 5253 |
| Db | 6770 | TCGCCGTTTACAATGTTCAATCTTTTCTAAAATGATATAGATAGTGTTCGCTCAAA | 6829 |
| Qy | 5254 | ATTTAAAAAATATAGAAATATTTATTCATAGATATCGCTCTGTAGATGATTTCTGCTAAAA | 5313 |
| Db | 6830 | CGTATTCAAATTTAGAAAATTTATTTTAGTTGATGATGCTCTACGGATACTTCTGCGGATA | 6889 |
| Qy | 5314 | TATGCAAGGAATATGAGAAAAAGATATAAAGAGTAAAAATTTTTTTTTCACATAATCATAGTG | 5373 |
| Db | 6890 | TCGTGACCTATTTTCAA AAAAGATAGACGTATATTTAGTTTTTTTCA TAAGAAATAATGGTG | 6949 |
| Qy | 5374 | GAGTATCAAATGCTAGAAATCATGGAATTAAGCGGAGTACAGCTGAATATATATATGTTTG | 5433 |
| Db | 6950 | GCTTATCTGATGCCAAAAAAAAGAAATATGATAGCTACTGGAGATTATTTATTTTAC | 7009 |
| Qy | 5434 | TTGACTCTGATGATGTTGTTGATAGATAGTATAGTAAAAAATATATTTTAAATATTTATAA | 5493 |
| Db | 7010 | TTGACTCTGATGATTTATTTATA--TAAAGAAATGCAATTTGAGGAAATGTTTGGATGTA | 7066 |
| Qy | 5494 | AAAGTAGAGTGAATTTATCTGGTTGTTGTGACCTACTTTTTCAGAAAAATATAAATAATTT | 5553 |
| Db | 7067 | GTAAGGGATATGATTCAGATATTTGTTTATGCTATGTTATGTTGAAGAAATAGGCAACGAA | 7126 |
| Qy | 5554 | TTCAAGTGAATAATCAAAATATTTGATTTTGAAGCAATTAATACCGTSCAGGACATGGGAG | 5613 |
| Db | 7127 | TTATAAATATGTTTTAGAGATGAACAATAGAAACAAATAGTCCGCTACAGCAATATC | 7186 |
| Qy | 5614 | AAAAAAATTTTATGAATTTGTATATAAATAATTTTTTCTACTCCTCTGTTTGTAAACTAT | 5673 |
| Db | 7187 | GAATAATTTATTAATTTATGATGACTATAAAGCTATTTTTTCTGTAGCACATAATAAATTTAT | 7246 |
| Qy | 5674 | ATAAGAAAAGATACATAACAGATCTTTTTCAAGAGAAATCAATGGTTAGGAGAGATTTTAC | 5733 |
| Db | 7247 | ATAAGAGAAATTTATTTCTACAAATGTGTTTACCAGTAGGAAAACTACACGAAATGAAT | 7306 |
| Qy | 5734 | TTTTTAACTGCAATTTATTTAAAGAAATATAGATAGAGTTAGTTATTTTGACTGAACATCTTT | 5793 |
| Db | 7307 | TTCTAACTTATAAATTTATTTGAAAGCAAAAAATAAATTTTTCTTTTCGATATAACACTT | 7366 |
| Qy | 5794 | ATTTTTATAGGAGAGGTATACTAAGTACAGTAAATCTTTTTTAAAGAGGTT | 5843 |

| | | | | |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------|------|-----------------|
| Db | 7367 | ATGTATATCGTATTAGACAGACACTATATGACTGTGTTCTTACAACTTT | 7416 | |
| RESULT 8 | AF349539 | 18239 bp | DNA | linear |
| LOCUS | AF349539 | | | BCT 05-APR-2001 |
| DEFINITION | Streptococcus agalactiae CNCTC 1/82 type V capsular polysaccharide synthesis operon, complete sequence. | | | |
| ACCESSION | AF349539 | | | |
| VERSION | AF349539.1 | GI:13549124 | | |
| KEYWORDS | | | | |
| SOURCE | Streptococcus agalactiae | | | |
| ORGANISM | Streptococcus agalactiae | | | |
| REFERENCE | 1 | (bases 1 to 18239) | | |
| AUTHORS | McKinnon, K., Chaffin, D.O. and Rubens, C.E. | | | |
| TITLE | Streptococcus agalactiae type V polysaccharide synthesis operon complete sequence | | | |
| JOURNAL | unpublished | | | |
| REFERENCE | 2 | (bases 1 to 18239) | | |
| AUTHORS | McKinnon, K., Chaffin, D.O. and Rubens, C.E. | | | |
| TITLE | Direct Submission | | | |
| JOURNAL | Submitted (16-FEB-2001) Department of Infectious Diseases, Immunology and Rheumatology, Children's Hospital and Regional Medical Center, 4800 Sand Point Way NE, Seattle, WA 98105, USA | | | |
| FEATURES | Location/Qualifiers | | | |
| source | 1. 18239 | | | |
| gene | /organism="Streptococcus agalactiae" | | | |
| -35_signal | /mol_type="genomic DNA" | | | |
| -10_signal | /strain="CNCTC 1/82" | | | |
| gene | /db_xref="taxon:1311" | | | |
| RBS | 792..18239 | | | |
| CDS | /gene="type V capsular polysaccharide synthesis operon" | | | |
| gene | 792..797 | | | |
| RBS | /gene="type V capsular polysaccharide synthesis operon" | | | |
| CDS | 818..823 | | | |
| gene | /gene="type V capsular polysaccharide synthesis operon" | | | |
| RBS | 910..2379 | | | |
| CDS | /gene="cpsA" | | | |
| gene | 910..915 | | | |
| RBS | /gene="cpsA" | | | |
| CDS | 922..2379 | | | |
| gene | /gene="cpsA" | | | |
| RBS | /note="similar to LytR" | | | |
| CDS | /codon_start=1 | | | |
| gene | /transl_table=11 | | | |
| RBS | /product="CpsA" | | | |
| CDS | /protein_id="AAK29647.1" | | | |
| gene | /db_xref="GI:13549125" | | | |
| RBS | /translation="WNSHRRHOKKHSHTPLRVNLFLLVIFILLVVSLFLMYRRHHP | | | |
| CDS | LAPHLNVIYGVIVLILVLSFLCIRKARIFPTIILVLASIFVATLYGPKSTIDL | | | |
| gene | TNNLKTSYSEIEMSVVPKDSKITNIEAVSKLAAPVKNDTSNITDLIEHKSEKGI | | | |
| RBS | SITPKQTSYDQYANRISQDSQAMVLNVAVSLIELSTPDPKSQIKTIYTKYKGI | | | |
| CDS | NRKNTGNGEVFNIVISGIDTFGISTVSRSDVNIIMVTNTHKVLITTPPRAYVK | | | |
| gene | IPDGGYDKLITAGLVGVETSMKLENLINDYVARINFSFLKLIIDLGAVTV | | | |
| RBS | YNDQAFTHKGNFPPVQVTLNSEQALGFVRVYSLGGDNRGNQEKVIAIINK | | | |
| CDS | LASQSVTKLNSISLQTSVQTNTIDNINLNQLSTGQRTFVESQALTHGSTG | | | |
| gene | ELPSYAMPQALYMSIDQSSLSNAKSIKNTMBE" | | | |
| RBS | 2372..3116 | | | |
| CDS | /gene="cpsB" | | | |
| gene | 2372..2377 | | | |
| RBS | /gene="cpsB" | | | |
| CDS | 2385..3116 | | | |
| gene | /gene="cpsB" | | | |
| RBS | /note="tyrosine phosphatase" | | | |
| CDS | /codon_start=1 | | | |
| gene | /transl_table=11 | | | |
| RBS | /product="CpsB" | | | |
| CDS | /protein_id="AAK29648.1" | | | |
| gene | /db_xref="GI:13549126" | | | |
| RBS | /translation="MKICLVSSSGHLAHLNLLKPIWEKEDRFWTFPKEDARSLIRE | | | |
| CDS | EIVYHCFPTNRNKNLVKTIAPKVRKERPDVIVISSGAAVAPFFYIKGLFCKT | | | |
| gene | VYIVFPRMDKPTLTGKLIVPYVTDKFIQVQEMKKVYPKAINLGIGP" | | | |
| | 6385..6891 | | | |

| | |
|------|------------------------------------------------------------|
| gene | KPALIEFSMOTSKDIHTALSNNVLMGLITPVVAHIERYNALENOKERVKEIINMGCTT |
| RBS | QINSHILKQKLEFKHFRKGRARYFLEENLVHFVADSMHNDLVRPPFLAEAYKIC |
| CDS | RDFGKERANQLFIENQAQSILKNHYI" |
| gene | 3115..3817 |
| RBS | /gene="cpsC" |
| CDS | 3115..3120 |
| gene | /gene="cpsC" |
| RBS | 3125..3817 |
| CDS | /gene="cpsC" |
| gene | /note="putative CPS chain length regulator" |
| RBS | /codon_start=1 |
| CDS | /transl_table=11 |
| gene | /product="CpsC" |
| RBS | /protein_id="AAK29649.1" |
| CDS | /db_xref="GI:13549127" |
| gene | /translation="MNKIANTVEINIFNLKKLWKKKFLITFVAIPATAGLIFYSLF |
| RBS | IVTPQYTSSTRIYVNPNTNNISITAOQLQAGSFLANDYKEIITSTDLKVISEKL |
| CDS | NYPSQLQKTIYSILKDTRVISLSEVDANPKMSOKLANSVREAAVSKIKAVTOVEDI |
| gene | TTLEKGNLKPAPSPFNKKNVLIGFIVAGLSTVIVIMGLDDRVRNTEEDIEKVLGL |
| RBS | TSLGIVPLNKL" |
| CDS | 3817..4526 |
| gene | /gene="cpsD" |
| RBS | 3817..3822 |
| CDS | /gene="cpsD" |
| gene | 3828..4526 |
| RBS | /gene="cpsD" |
| CDS | /note="CPS chain length regulator/exporter" |
| gene | /codon_start=1 |
| RBS | /transl_table=11 |
| CDS | /product="CpsD" |
| gene | /protein_id="AAK29650.1" |
| RBS | /db_xref="GI:13549128" |
| CDS | /translation="WTRLEIVDSKLROAKTEEYFNARTNIQFSGKENKILAITSVR |
| gene | EGEKSASTSLALSQAQFKTLIDADTRNSVMSGTFKATGIKGLTNLNSGNADL |
| RBS | GDIICTENPRLVTVPSGKVPNPPTALLQNAVFNKLEAKNIIDYIILDPPIGLV |
| CDS | DDAILNACDGFILVTQAGRIKRVYKAKEQESQSKFLGILNKNVSNVATYGDY |
| gene | GDYNGYKGRDRKRK" |
| RBS | 4564..5927 |
| CDS | /gene="cpsE" |
| gene | 4564..4569 |
| RBS | /gene="cpsE" |
| CDS | 4578..5927 |
| gene | /gene="cpsE" |
| RBS | /note="glucose-1-phosphate transferase" |
| CDS | /codon_start=1 |
| gene | /transl_table=11 |
| RBS | /product="CpsE" |
| CDS | /protein_id="AAK29651.1" |
| gene | /db_xref="GI:13549129" |
| RBS | /translation="MIQTVVVYFSAULTLITPNFKSNKOLLFVLLHYIVFVLSDF |
| CDS | YRDFWSRGYLEEFKWLKYSFYIFISSLPFIKNSFTTRLSGFTFIAMNSILLYL |
| gene | LNSELYRYKYVYAKFSRDTKVLITNKDSLKMTFRNKYDHNVIACVILDSKDCY |
| RBS | DLNLSRIRINKDALTSELTCITVDQAFINIPBLFKYQIQDIINDIEMGVIVNN |
| CDS | VEALSFNIGEKRIQTEGYSVITYSMKFKYSHLIAKRLDITGAILIGLIGLVAI |
| gene | FLVQPIKDGQPAIFSQNRVGRNGRIIFRYFKPSRMVDAEQIKDILLVHNMQTMGLMF |
| RBS | LDDPRTIKTGKFKRTSIDELPOFYNVLKGDMSLVGTRPTVDEYEKYNSTQKRLS |
| CDS | FKPGITGLWISGRNNTDFDEIVKLDVQYINENSIVSDIKIILLTLKVLVLTGAK" |
| gene | 5942..6400 |
| RBS | /gene="cpsF" |
| CDS | 5942..5947 |
| gene | /gene="cpsF" |
| RBS | 5951..6400 |
| CDS | /gene="cpsF" |
| gene | /note="beta-1,4-galactosyltransferase enhancer" |
| RBS | /codon_start=1 |
| CDS | /transl_table=11 |
| gene | /product="CpsF" |
| RBS | /protein_id="AAK29652.1" |
| CDS | /db_xref="GI:13549130" |
| gene | /translation="MKICLVSSSGHLAHLNLLKPIWEKEDRFWTFPKEDARSLIRE |
| RBS | EIVYHCFPTNRNKNLVKTIAPKVRKERPDVIVISSGAAVAPFFYIKGLFCKT |
| CDS | VYIVFPRMDKPTLTGKLIVPYVTDKFIQVQEMKKVYPKAINLGIGP" |
| gene | 6385..6891 |

| | | |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|
| RBS | /gene="cpsVG" | |
| | 6385..6390 | |
| CDS | /gene="cpsVG" | |
| | 6400..6891 | |
| | /gene="cpsVG" | |
| | /note="beta-1,4-galactosyltransferase" | |
| | /codon_start=1 | |
| | /transl_table=11 | |
| | /product="CpsVG" | |
| | /protein_id="AAK29653.1" | |
| | /db_xref="GI:13549131" | |
| gene | ONCOWSKFSLSDDDMNSYMKAEIIVTHGGPATFNVAYSKGKKTIVVPRQOPGSHVNN HOYDFVNKVTMYNFDIVDIERLQNVVYEGWMRPFEITNSSNFIEEFKVLKELCD ENQ | |
| | 6866..7972 | |
| RBS | /gene="cpsVH" | |
| | 6866..6871 | |
| CDS | /gene="cpsVH" | |
| | 6878..7972 | |
| | /gene="cpsVH" | |
| | /note="putative repeating unit polymerase" | |
| | /codon_start=1 | |
| | /transl_table=11 | |
| | /product="CpsVH" | |
| | /protein_id="AAK29654.1" | |
| | /db_xref="GI:13549132" | |
| gene | /translation="MKINKSLFYIAIFLVNPFKSLGLGEGNSAYKIIVMLVAILLGGI KFLDLSLYFERRKLOVIFLFIATILNLFVHKVTFILFIILFALKDILSKKAFSII IGSRILGVNKLVIKLVNPFVROQOFLRSDLGFGHPNFTHNFALTIPLY IVLVNKLKPVWMLFILTANLYQVTFSTRGYVILFVILVIVTKNSLKRVMKL APYQFLLVPFPLSSIFFNNSFVQKLDVLLGRHYAHLQLVGLTPFGNSFKETS VLPDMSYMLSLMSLVGVVTFMCFMIIYYSKIIIIIEQLLLLFIIMSILILFTSFYPSV VMNLSWIVFGICDGEPIKKEFTIVNNI" | |
| | 8001..8736 | |
| RBS | /gene="cpsVM" | |
| | 8001..8006 | |
| CDS | /gene="cpsVM" | |
| | 8011..8736 | |
| | /gene="cpsVM" | |
| | /note="putative glycosyltransferase" | |
| | /codon_start=1 | |
| | /transl_table=11 | |
| | /product="CpsVM" | |
| | /protein_id="AAK29655.1" | |
| | /db_xref="GI:13549133" | |
| gene | /translation="MIPKVIHYCFWGGNPLDNLLKKYIKTWREQCPDYIEIWNHNY DVSXNVFWREAYTKKNFAYSDYARLDIIYTYGGFYLDVDELKLSLDPLRIHECFPLA REISCDVNTGLIIGAVKGGHPLKSNMSIYDKSLDLSLTKTCVEVTTNLLINRGLKKN IIQKIDITIIIPRYFNPKNLLTGKVDCLISVTVSIHHYEGSWKSSSFISDSLKIRV LIIDFLFGYGYTFMRLRLPLKUK" | |
| | 8754..9649 | |
| RBS | /gene="cpsVN" | |
| | 8754..8759 | |
| CDS | /gene="cpsVN" | |
| | 8762..9649 | |
| | /gene="cpsVN" | |
| | /note="putative glycosyltransferase" | |
| | /codon_start=1 | |
| Query Match | 11.4%; | Score 798.8; DB 1; Length 18239; |
| Best Local Similarity | 48.2%; | Pred. No. 2.6e-81; |
| Matches 3323; Conservative | 0; | Mismatches 3382; Indels 185; Gaps 31; |
| Qy | 16 | TGGCAATATTTCATATGATAGTAGGTTGCAATTTCTGCAATCTTAAACAGTCATATACCAA 75 |
| Db | 4573 | TAGCGATGATTCACACAGTCTGGTTTATTTCTGCAAGTTTGACATTAACATTAATTA 4632 |
| Qy | 76 | ATGCTGATTTAAATCG----TTCTGGAATTTTATCATATGATGGTTCATTTTTCAT 132 |
| Db | 4633 | CTCCCAATTTTAAAGCAATAAGATTTATTTGTTGTTTCTATTCATATATATATATCTCT 4692 |
| Qy | 133 | TTTTTATATCTCGTAGCCAGTTGAAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAAA 192 |

| | | |
|----|------|------------------------------------------------------------------------|
| Db | 4693 | TTTTATCTTCTGATTTTTCAGAGACTTTTGGAGTCGTGCTATCTTGAAGAGTTTAAAA 4752 |
| Qy | 193 | AAACATTTAACTATAGTATAATAATTTTGCAAATTTTCTTACGGCAGTATCATTTTGTGG 252 |
| Db | 4753 | TGCTATTGAAATACAGCTTTTACTATATTTTCATATCAAGTTCATTTATTTTATTTTAA 4812 |
| Qy | 253 | AGAATAATTTGCGCACITTTCAAGACGGTGGTCCCGTGATTTTTCACATTAATAACTTCGTT 312 |
| Db | 4813 | AAAACCTCTTTTACAACGACACGACTTTCCTTTTACTTTTATTTGCTATGAATTCGAATTT 4872 |
| Qy | 313 | TGCTATACCTTATTAACTGAATTTAAGCAGTTTAAAGGATAGCTTCTATTTTTCGACAA 372 |
| Db | 4873 | TATATATCTATTGAATTCATTTTAAATAATTTATCGAAATATTTCTTACGCTAAGTTT 4932 |
| Qy | 373 | TCATCAAAAAAGAGAGATTCTAATTAACAACGGCTGAACGATGGGAAAAATATGCAAGTTT 432 |
| Db | 4933 | CACGAGATACCAAGTTGTTTGTATAACGAATAAGGATTTCTTTATCAAAAAATGACCTTTA 4992 |
| Qy | 433 | TATTTGAAATCATAAACAATAATTCAAAAAATCTTGTTCATTTGGTAGTTTTAGGTACAG 492 |
| Db | 4993 | GGAATAAAT---ACGACCAATAATTATATCGCTGTCTGTATCTTTGGATTCCTCTGAAAAGG 5049 |
| Qy | 493 | AAATAGATAAAATTAATTTATCATTTACCGCTCTATTATTCTGTGGAAGAAGCTATAGAGT 552 |
| Db | 5050 | ATTGTTATGATTTGAAACATAACTCGTTAAGGATAATAACAAGAAGTCTCTTACTTTCAG 5109 |
| Qy | 553 | TTTCAACAAGGGAGTGGTCGACCAAGTCTTTTATAAATCTACCAAGTGAGTTTTTAGACG 612 |
| Db | 5110 | AGTTAACCTGCTTAACTGTTGATCAAGCTTTTATTAACATACCACTTGAATTTATTTGGTA 5169 |
| Qy | 613 | TAAGCAATTC-----GTTTCAGATTTTTCAGTTTGGTAGTTTGGTATTTGATGTAAGCCGTTG 663 |
| Db | 5170 | AATACCAAAATACAAGATATTATTATTAATGACATTAAGCAATGGAAGGAGTATTTGTCATGTTA 5259 |
| Qy | 664 | ATATTAAATTCATTTCCGTTTACTGCGTTGAAAAAACAATAAATTCACACTGCTAGGTGACC 723 |
| Db | 5230 | ATGTAGAGCACTTAGCTTTGATAATATAGGAGAAAAGCGAATCCAAACTTTTGAAGGAT 5289 |
| Qy | 724 | ATAGCAATTTGTAATTTTTCACAAATTTTATAAGCCTAGTCATATCATGATGAACGAC 783 |
| Db | 5290 | ATAGTGTATTACATATCTTATGAAATTTCTATAAATATAGTACCTTATAGCAAAACGAT 5349 |
| Qy | 784 | TTTTGATATATCTCGAGCGGTAGTCGGGTAAATTTTGTGGTATAGTTTCTATTTTGT 843 |
| Db | 5350 | TTTTGGATATCACGGGTGCTATTATAGGTTTGTCTCATATGTGCAATTTGTGCAATTTTTC 5409 |
| Qy | 844 | TAGTTCCAAATTTATCGTAGAGATGGTGCACCGGCTATTTTGTCTCAGAAAACGAGTTCGAC 903 |
| Db | 5410 | TAGTTCCAAATCAGAAAAGATGGTGCACCGGCTATCTTTTCTCAAAATAGAGTAGGTCTC 5469 |
| Qy | 904 | AGAATGGACGCATATTATACATTTCTACAAGTTTTCGATCGATGTATGTTGATGCTGAGAGC 963 |
| Db | 5470 | GTAATGGTAGGATTTTATAGATTCTATAAATTCAGATCAATTCGAGTAGATGCGAACAAA 5529 |
| Qy | 964 | GCAAAAAGACTTCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTTAAATGGGAA 1023 |
| Db | 5530 | TTAAGAAAAGATTTATTAGTTTCACAAATCAAAATGACAGGG-----CTAATGTTTAAAGTTAGA 5584 |
| Qy | 1024 | AAACGATCTTAGAATTTACTCCAATTTGACATTTTCATACGCAAAAAACAAGTTTACAGCAGT 1083 |
| Db | 5585 | CGATGATCTTAGAATTTACTAAAATAGGAAAATTTATTCG-AAAAAAGAGATAGTAGT 5643 |
| Qy | 1084 | TACCACAGTTTATAATGTTTTTAAATTTGGCGATATGAGTCTAGTTGGTACAGTCCACCTA 1143 |
| Db | 5644 | TGCTCAATTTCTATAATGTTTTTAAAGGGTGATATGAGTTTAGTAGGAACACGCCCTCCCA 5703 |
| Qy | 1144 | CAGTTGATGAAATTTGAAAAATATCTCTCGGTCAAAGAGACGATTCAGTTTTHAACCCAG 1203 |
| Db | 5704 | CAGTTGATGAAATGAAAAGTAAATTTCAACGCAAGACGACGCCCTTAGTTTTHAAGCCAG 5763 |
| Qy | 1204 | GGATTACAGGTCTCTGCGAGGTTAGTGTCTGTAGTATATACAGACCTTCGACACCGTAG 1263 |

Db 5764 GAATCACTGGTTGTGGCAATATCTCGTAGAATAATATTAATCTGATTTTGAAGAAATCG 5823
Qy 1264 TTCCGTTGGACCTTAGCATACATTTGTAATTTGGACTATCTGCTGAGATATTAATAATTTTAT 1323
Db 5824 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTAAGATTTATC 5883
Qy 1324 TAAGACAGCTGAAGTTGTTATTTGAGAGAGGAGTAAGTAAAGTAT 1373
Db 5884 TCCTAACACTAAGGTTAGTTTACTCGGGACAGGAGCTAAGTAAAGGTTTGAAG 5943
Qy 1374 -----ATGAAAGTTTGTGGTCGGTTCTTCAGGGGACATTTGACTCACTTCTGTTAT 1426
Db 5944 GAATATATGAATTTGTCTGGTTGGTTCAAGTGGTGTCTATCTAGCACACTTGAACCT 6003
Qy 1427 GTTAAACCGTTTGGAAAGGAAGAACGTTTGGGTAAACATTTGATAAGAGGATGC 1486
Db 6004 TTTGAAACCCATTTGGGAAAAGAGATAGGTTTGGGTAACTTTGATAAAGAGATGC 6063
Qy 1487 AAGAAGCTTTTGAAGAAATGAATAATGTTATCCATGTTTACTTTCACAAATTCGAATCT 1546
Db 6064 TAGGAGTATTTCTAAGAGAAGAGATGTTATCAATTTGCTTCTTCCAAACCCGTAATGT 6123
Qy 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTCAAAATTTTACGTGATGAGAACCCAGA 1606
Db 6124 CAAAACCTTTGGTAAAAAATCTATTTCTAGCTTTTAAAGTCTCTTGAAGAAAGAACCCAGA 6183
Qy 1607 TGTTATTTATTCATCTGGTGGCGGCTTGTCTCCCTTCTTTTACATCGGAAACTATT 1666
Db 6184 TGTTATCATATCACTGTTGGCTGCTGAGCAGTACCATTTTATATTTGGTAACTTAT 6243
Qy 1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTAATAATCTACATTAACCTGG 1726
Db 6244 TGGTTGTGAAGACCGTTTACATAGAGGTTTTCGACAGGATGGATAAACCAACTTTGACAGG 6303
Qy 1727 AAACTAGTTATCCCGTAAACAGATTTTATTTGTTCCAGTGGGAGAAATCAAGAGGT 1786
Db 6304 AAAATTAGTGATCTCTGAACAGATAAATTTATTTGTTCCAGTGGGAGAAATGAAGAAAGT 6363
Qy 1787 ATATCCTAAATCTATTAACCTTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAT 1846
Db 6364 TTATCCTAAGGCAATTAATTTAGGAGGAAATTTTAAATGATTTTGTACAGTGGGACA 6423
Qy 1847 CATGAACACAGTTTAAATCGATTTGATAAAGAGATTTGATTTATGAAAAAATAATGGAAGT 1906
Db 6424 CATGAACACAGTTTCAACCGTCTTATTAAGAAAGTTGATAGATTAAGAGGACAGGTCT 6483
Qy 1907 ATACCCGAGAAATATTTATTCACACAGATATCTGACTATATTTCCAGATATTTGCAAG 1966
Db 6484 ATTGATCAAGAGTGTTCATTTCAACCGGTTTACTCAGACCTTTGAAACCTCAGAAATTTGCAG 6543
Qy 1967 TATRAAAAATTTCTCAGTTTACAAAGAAATGGAACAATATATTAACAATCAGAAGTAGTT 2026
Db 6544 TGGTCAAAATTTCTCTATATGATATGATGAATCTTATCATGAAGAAGCTGAGATTTGT 6603
Qy 2027 ATTTGCCAGGAGGCCCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAAACAATTA 2086
Db 6604 ATCACACATGGCGGCCAGCGAGCTTTATGAATGAGTTCTTAAAGGAAAAAACAAT 6663
Qy 2087 TTGTTTCTAGACAAAAAAGTATGGTGAAACATGTAATGATCATCAAGTAGAGTTTGTGA 2146
Db 6664 GTGGTTCTTAGACAAGAACAGTTTGGAGAGCATGTGAATATCATCAGTGGACCTTGT 6723
Qy 2147 AGAAGAAATTTACAGATAATTAATTTTATTTATAGAAATATAGATGATTTGTTGAA 2206
Db 6724 AATAAGGTAAGAAACAAATGTATTAATTTGATATCTGTTGATATTTGAAGGTTTACAAAT 6783
Qy 2207 AAAATTTATGAGTTTCTTAAGCAAACTAACTTTTACATCAATAATAATTTT 2263
Db 6784 GTAGTCTATAGGGAATGATGAATCGTCCGTTTGTAGAACTAATAGTAGTAATTTAT 6843
Qy 2264 GAAGATTAAAACAAAATAGTTGAAAAATTTTAAATGAGGATCAAGAAAAATGAATAAAAA 2323
Db 6844 GAAGAAATTTAAGGTAATTTAAGAGGTTGTGCGATGAATAATCAATAAAAACTTTTAT 6903

Qy 2324 AGATGCAATTTTGTATAATGGCTTATCATAAATTTTCTCAGATTTTACTGGAGAGGATAC 2383
Db 6904 TTATATTGCAATAATTTTAGTTAAATTTTAAATCACTGGGTTTAGCGAGGGAACCTC 6963
Qy 2384 AGATTTATCATCTCTCTCAGGAGATGCACCAATTTAGTTCTCTCAGATACCTGTAT 2443
Db 6964 AGCTTACAAAATAGTG-----ATGTTAGTTGCA 6992
Qy 2444 AATTAATTTAAATTTCTCAGGATTTATATGTTGAAATTTTACAAAAGATGAGCAAAAATAT 2503
Db 6993 TTTTACTGTGGGAATAAAATTTTATTAGATAGCTTTTATTTTGAAGAGAGAACTCG 7052
Qy 2504 AAAGAAATAGGATATATGAACGAGTTAAATGTTTACAGATTTTCTCTAATATATCAGAA 2563
Db 7053 TGATCATCTTTTATATTTATTCGCGACCAATTTTGAATTTATTTCT----- 7097
Qy 2564 AAACTATTTGATATGATCTGTTTAGATTTTATTTAGAAATGATCTGAGCTTTTGAATAC 2623
Db 7098 ----TTGTTCTATAAGGTTACTTTTTATATACTTTTAAATTTTTTTTCTAGCATTAAGGAC 7153
Qy 2624 TATTTACAAAGATTTGTTTATTTGATAGAATAAAAAACATGGTCTAAGAAATAAGATTTG 2683
Db 7154 ATCTCTCTAAAAAAGCTTTCTCTATATAATAGATCGCGTATTTTGGGAGTTCTATTA 7213
Qy 2684 GTTCTAATTTGGGTTTTCGCTCCACATGATTTTGTGGCAATTTCTTTTATCAAAATGAAAAAG 2743
Db 7214 AATCAAAATTTTGTGAAATTAGATTTAATAGAAATTAAGTATGTCAATTTTATAGGAT 7273
Qy 2744 AAACAGCTTATTTATTTAAGTAATCTAAATGTCAGATGAACATTTTATACAGCAATTA 2803
Db 7274 GGACAAATTTTCTGAGAAGTGA-CTTAGGTTTGTGTCATCTTAACCTTTATTTCAATATTT 7332
Qy 2804 TAGAAAAATATGAATTTTCAAAATAGATTTCTAAATATGAA-----ATTAAAGATATATA 2860
Db 7333 TTTTGTCTAACTATTTTCTGTATATTTGTACTCAATTTATAACGACTAAAGCCTGTGT 7392
Qy 2861 AGTGGAAAAATCAACATCTTCTCTATGTTCTTTTACAGATGATTTCTATTTGATGAATTC 2920
Db 7393 GATGTTTATTTTAAACATTAATAATTTATTTGTACCAATATCTTTTCAAGGACAGG 7452
Qy 2921 TAAATGCAGAAATTTAGGTTTATTTGCTAGAAAGTTAAATAAGAAATAAGAAATAATCTA 2980
Db 7453 GTATATATCTGAATTTTATTTATTTGTTACTCATTTATGTCAGAAAGATAGCTTAAATA 7512
Qy 2981 AA----TTTAAAGAAATTTACTAAAAATAAAATAGTTGATTTGTTGAGAGTAATGTA 3036
Db 7513 AAGAGTATTTAAGAAATTTAGCACCTATGTACAATTTTTTTTATTTAGTATTTTACCTTTT 7572
Qy 3037 TGTTTAAATTTTAAATATGACCCGGAATTTTATTTTAAAGTACTTTCTGGTTGATTA 3096
Db 7573 GAGTTCTACAATTTTAAATTTTAAATTTTGTTCAAAAATTAGATGTTCTTTTAAACAGG 7632
Qy 3097 TTTTATTTCCAGAGAAAAAGTATGTAATTTTAAATTTTATTTTATGAAATTTAATTTTATTT 3156
Db 7633 TAGATACACATGCTCATTTTACAACTTTGATAGTTTAACTCTCTTTTGGAAAAATGTTT 7692
Qy 3157 ATATAAATTTTGAAGAACTAAGCTTAATTTTAAAAATGAAATTTTATTTGTTTATTTAT 3216
Db 7693 TAAGGAAACAAAGTGTCTTATTTGATAATAGCTACTATGTTTATTTGAGTATGTATGGTGT 7752
Qy 3217 GGTCTATATTTATTTTGTGTTTTCAGTAGTCAAGATGATGTTGTTGTAATAAATTTTGA 3276
Db 7753 AGTACTTACCATGTTTGTATGATTAATCTATTTATCTATAGTAAAAAGA-----TAA 7805
Qy 3277 GATTTTTCAGATTTTACTCTCCCAATTTTGGATTTTGGATTTTGCATTAATGTTATTAAT 3336
Db 7806 TCATAATTTGAACCTCAACTACTCTCTATTTATAATGTCATAATATTTTACTGAAAGTT 7865
Qy 3337 TGTATTCATTTATAATTTGATTAATAAAAAATTTAAAAAATAGTATCTTTTGTAGTTT 3396
Db 7866 TTTATCCAGTGTGGTAAATGAATATTTAGTTGGCTAGTTTGTGGTAAAAATATTTTGTGATG 7925

| | | | |
|----|------|-------------------------------------------------------------------|------|
| Qy | 3397 | TAGTTTTATTAGGTATATCTGCATGTGTATATTATTCAAATGGGAAAGATATTGTATTTTT | 3455 |
| Db | 7926 | GTATCGAACCTTATAAAAAGGAAATTTACTATTGTGGAATAAT--ATATGACATATTGGCTC | 7983 |
| Qy | 3457 | TAGACACACACCTTATAGGACTAGACTATTCTTTAAACAGGCGTCAAAACAAGGTTGGTTG | 3516 |
| Db | 7984 | TGATATGGCAGGAGGTAAAGGAATAATGATACCTAAAGTTATACATTAATTGTGTGGTTT | 8043 |
| Qy | 3517 | GCTTTATGAACATCTCTCAGGTTAAATACCACATCAAAATTATAGTTTCAATTCCTCGTTAACTC | 3576 |
| Db | 8044 | GGAGGAATCCCTTACCAGATNAATTTAAAGAAATATATAAAACTTCGAGAGAACATGT | 8103 |
| Qy | 3577 | TTGCACATTATAAAAAATAAATGCACAAATTTTTTTCTTGTGTCTTGTCTTTTATACCGA | 3636 |
| Db | 8104 | CCGGATTATGAAATTAATTGAATGAAATGAGCATAAATTTATGATGT-----8147 | |
| Qy | 3637 | TCATTTTAAGTGGATCGAGAAATGGTAGTTTATCGCTAGCAATATTAATTTATATGCTTGT | 3696 |
| Db | 8148 | --TAGTAAAAATGTTTTTATAGAGAGAAGCATATACTAAGAAGAAATTTGCTTATGTTTCT | 8205 |
| Qy | 3697 | TATGGAGATATATAGG--TGGAAAAATTTGCTTGGATATAAAAAAGCTAAATAGTAAATTTGTA | 3755 |
| Db | 8206 | GACTATGCAAGATTGGAATATTATTATCTATTATGGGGGTTCTATCTAGATATCTGATGTCG | 8265 |
| Qy | 3756 | ATACTACTTATTTTAAATACATGAATTCGCTTTACCATGAAATTTTGGCTGTTTATAAAT | 3815 |
| Db | 8266 | GAGCTTTTAAAAAGTTTTAGATCCTTTCGAGGATTCATGAGTGTTTCTTAGCAAGGAGATT | 8325 |
| Qy | 3816 | TCTAGAGAAATCAAGTAACGAAGCTAGATTTTATTTTATCAAGGAAGTATTGATAAGTA | 3875 |
| Db | 8326 | AGTTGTGATGTGAATACAGGATTAATTAATTTGGCGCTGTTAAAGGACATCACTTTTAAAA | 8385 |
| Qy | 3876 | TTAGAAAAACAATATTTTATTTGGATATGGAATATCCGAATATTTCAGTTACGGGAACCTGG | 3935 |
| Db | 8386 | TCAAAATATGCTATATATGACAAAGTGATTTAACTTCTCTTAATAAGACATGTGTAGAG | 8445 |
| Qy | 3936 | CTCGGAAGTCAATTCAGGCTATATATCAATTTTTTTTATAAAATCAGGAATAGTTGGGTGATT | 3995 |
| Db | 8446 | GTTACAACATAATTTATTGATAAACAAGGGCTTAAAGAT--AAGAATATTATTCAAAGA | 8503 |
| Qy | 3996 | TTACTGATGTTTTCTTTTTTTTATGTTATAAAAAAGTTATGGAGTTAATGGGGAACA | 4055 |
| Db | 8504 | TTGATGATATAACAAATATATCCGAGAAATATTTTTAATCCAAAGAAATTTATTAACAGGTA | 8563 |
| Qy | 4056 | GCATAATTTTATTTTATCATCATATAGCCATATTTTTTCATATATGAAACAATAGATCCGATT | 4115 |
| Db | 8564 | AGGTTGATGTCTGACTAGTGTGTACCTATTCTATACATCAATACGAAGGAAGTTGGAAAA | 8623 |
| Qy | 4116 | ATTATTATATTAGTACTATTCTTTTCTTCAATAGTATTTTGGAAATAATAAAATTTTAAA | 4175 |
| Db | 8624 | GTTCTTCATTATTTTCAGATTCTCTAAAGATTAGAGTAAGGCTCATAAATTTGAATTTTAT | 8683 |
| Qy | 4176 | AAGGATATGGAGACAAAAATGAATGATTTAAATTTTCAGTTATTGTACCAATTTATTAATGT | 4235 |
| Db | 8684 | TTGATATGGTACTTTATAGAAATGCTTCTTAAGGTTTTCTAAAGTTTAAAGAAATAGTTATTAT | 8743 |
| Qy | 4236 | CCAGATTATCTTGATAAATGCTATTACAGTATTATTATAACCAACATATACTAATTTTAGA | 4295 |
| Db | 8744 | TGTAATAGAAGGACGAATGTTATATATTTTACTCCTCATTTTACAGAGGCTACCGA | 8803 |
| Qy | 4296 | GGTTATTCTCTTAATGATGGAAGTACTGATGTTCTTGAGAAAAATTTGCTTTAAACTATAT | 4355 |
| Db | 8804 | TTATCCTATC--TGTAAGATTCTCTATGTAATCAGACAAATATAAAATTTTATATGGTTGAT | 8862 |
| Qy | 4356 | GAAGAACGATGGAGAAATTAATATTTACAAGAAAAATTAATGGCGGCTTAGCAGATGCTCG | 4415 |
| Db | 8863 | TGTTGATGATGGTTCTCGGAGACAGTACTAAGGAAATTTGTAAGTAATTTATATAAAAGAA | 8922 |
| Qy | 4416 | AAATTTTCGGACTAGAAACATGCACAGGTAAATATATTGCTTTTTTGTCCGATTTCTGATGACTA | 4475 |
| Db | 8923 | TAAAGTTAGTATTCTGATATCTGTA--TAAACGTAATGGCGGTAAGCAATTCAGCCTATAA | 8979 |
| Qy | 4476 | TATAGAAGTTGCAATGTTTCGAGAGAAATGCATGATAATATAACTGAGTATTAATGCCGATAT | 4535 |

[illegible]


```
ISKFAPKAYOKATTGADSOLEMTKTELELSFEYLEMRDYAISKGVETESTPDBES
LEFLISTDMFIYKIPSEITNLPYLEKIGKQKVKVILSTGMVNEEIHQAVNLRQNG
TDEIHLCTTEYTPPTPPKSNLNLNVIHLTKDFKDLTIIGSDHSISGSEVPAAAMAGVA
VIEKFTLVDMEGPDKASATPDILAAALVGVRIVEQALGRFEKIPDPVEERKQIVA
RKSVALVKPKKGIKDIYSIENTVTRPGNGISPMNWDILGQEAQDDFEDEVIRDSRF
ENOLPEL"
complement (1094. .2494)
/gene="cpsL"
/note="synonym: SAG1162"
complement (1094. .2494)
/gene="cpsL"
/note="identified by match to PFAM protein family HMM
PF01943"
/codon_start=1
/transl_table=11
/product="polysaccharide biosynthesis protein CpsL"
/protein_id="AAN00044.1"
/db_xref="GI:22534191"
/translation="MNTKKLLQSGFIYTLGNLLVQGLAFITLPIYTRVISTEVYGVQS
LYVAMGIIITLFIQIGTSGSLSSARVKYIGEEFKSYGSFASVGNINFLIIVLLIAFLPR
SFLAPLVGESIFELLMVQSYANVYVTFQGYPIQQORSLANLILSLANAVSVALS
LFLTFHNSDDFLSRVFGAFVPTIITGIVAFAYIYTHSKSPNPKYFRFIVTVSIPLIIF
HLGHQNLNOLDRLMARLYSTKEVAMYSFGSLGMIQIVLNSINMAWIPWFFDARK
EKLQLSTYSRYLVGLVETLGVLTVPESLAQIMGGDKYSSVQSFSLIIVSYELVF
LYTFPVMIQFYANTWPIPIGTLLAAGVNMLLNLVLIPIHYAAGAAMATIIISYALLI
FHHIVSKYKHYSDVRQYIILSGIVFSYAMLMNMFLGNIVIRWSLGIILILIVISIV
FQKVIDLLGKKRRR"
complement (2491. .3447)
/gene="cpsK"
/note="synonym: SAG1163"
complement (2491. .3447)
/gene="cpsK"
/note="similar to GP:13549137; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polysaccharide biosynthesis protein CpsK(V)"
/protein_id="AAN00045.1"
/db_xref="GI:22534192"
/translation="MNNFVCHTLVHLITIIKLKDKENTRIFIGDTITNTYTWKTIN
BOGRTBAVKEFAVREQLOKNIIEVMELVDSLNLNHYPERVDVQVYLFNDDTLIGRYM
VVLGKNHLLLEDGVCQAKLFLGGSVVKVITKYLEKYPVGFESKYCLISIVNSIV
GLPHDIRSKYKELPRKLFDSLNKEOKSLIFKIFTKPLTIIPKSVLLTQPLAQDK
GLYPTTRFQSIQDQYFDIDIVQEXYETLGNVYLVKHPRDVVDVSKLPVELLPSNIP
MEITELMLTGRFEPFGITHSTALDFLTCDVKKTLVDLKDIK"
complement (3444. .4409)
/gene="cpsJ"
/note="synonym: SAG1164"
complement (3444. .4409)
/gene="cpsJ"
/note="identified by match to PFAM protein family HMM
PF00535"
/codon_start=1
/transl_table=11
/product="glycosyl transferase CpsJ(V)"
/protein_id="AAN00046.1"
/db_xref="GI:22534193"
/translation="MDKVSIIIPVYNVQSFINECIESVLAQTSYNLBIILVNDGSTDN
SGDTCVYSEIDGRIFVFKHNGSLDARNYGISRATGDVYIILDSDDLYLKDAIER
MVPESEKNSRIVLGCYVKEQEHINIVLEDEMLETISPQVATQNIYNYVDAYRAJIT
VAHNKLKRELFTSLCTCPVGKLHDEDFLTLYKLAKNIIIFFRNTYAIRRENSIMT
GSYNIKRHVAELKREYLLEKYPDLVQFOSERALIKTMEVNLLIELYNNKPNFYKFFHTL
KTEYKKTIFDPKRRKMLLKIKYLYKVCVHFPTLNCRRKKKINK"
complement (4402. .5385)
/gene="cpsO"
/note="synonym: SAG1165"
complement (4402. .5385)
/gene="cpsO"
/note="identified by match to PFAM protein family HMM
PF00535"
/codon_start=1
/transl_table=11
/product="glycosyl transferase CpsO(V)"
```

```
/protein_id="AAN00047.1"
/db_xref="GI:22534194"
/translation="MELISIIIPVYNGBIYIGRCLDSILSEQTYQNLEIILIDGSSDR
TGDCIKCYLEDRRIKYFOENRQSVARNNGVLRCTGDMWIAFLDSDVLYPYSIEVM
YIKYQATNADIVLTSIGNFNNTYNTINSQYKEIKLYLTLEVALEEMYYGKTYGVSPF
AKIYPRSNLISNTPYRGGKTHEDMTTKLISCASKIACVDIVTAVVVSNDSTRTKF
NERMLYFPEAIQNNIVFINLPHNTSLISAVIYNEVFGGIDICGKMDPKLYDTVDY
YRKKYKYPKTLIFNRRISVGEKVKYILFISSIRYFTIVRKIYNLRLSRNG"
complement (5389. .6276)
/gene="cpsN"
/note="synonym: SAG1166"
complement (5389. .6276)
/gene="cpsN"
/note="identified by match to PFAM protein family HMM
PF00535"
/codon_start=1
/transl_table=11
/product="glycosyl transferase CpsN(V)"
/protein_id="AAN00048.1"
/db_xref="GI:22534195"
/translation="MLYIPTPTFNRGYRLSYLSDLCNQTNKNFIWLIVDDGSDSTK
EIVSNYIKENKVISIVLYKRNKGKSHAYNLAMRYMQPSDYHVCVDSDDMLLEDAVEII
FKDLESLTISNRVGLVPRYSLNQGNLNPRLKILEVNIPLDKYKHLKTCICVINN
AYLVDPEFCPEGENFYSBEIMYIYLSKGYFCQPNRKLYCFDYLEDGLTSNIFKLWR
KNFKGTFISLENSYMYVMSFPNIFDRWNSAIKIKMNIQALKMTTLGVIPFLKSEAGW
KILLGFSYLKWKARFKKSE"
complement (6302. .7027)
/gene="cpsM"
/note="synonym: SAG1167"
complement (6302. .7027)
/gene="cpsM"
/note="similar to GP:13549133; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polysaccharide biosynthesis protein CpsM(V)"
/protein_id="AAN00049.1"
/db_xref="GI:22534196"
/translation="MIPKVIHYCWFQGNPLPDNLKKYIKTWBQCPDYEIEIWNENHY
DYSKNVMEAYTKNFAYVSDYARDLIITYYGGFYLDITDVELLKSLDPLRIHECFLA
RETCDVNTGLIIGAVKGHFLKSNMSIVDKSDLTSLNKTCEVNTNLLNRGLKNQK
LIQKIDITIPENYENKNLLTGKVDCLTSVTYSIHVYEGSKSSFSIDSUKIRV
LIIDFLGQVTRMLLRFLKXK"
complement (7066. .8160)
/gene="cpsH"
/note="synonym: SAG1168"
complement (7066. .8160)
/gene="cpsH"
/note="similar to GP:13549132; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polysaccharide biosynthesis protein cpsH(V)"
/protein_id="AAN00050.1"
/db_xref="GI:22534197"
/translation="MKINKNSLPYIAFLVNFPSLGLBGENSAYKIVMLVAILCGII
KFLDLSYFERRKLVIFILLFIATILNLPVHKVTFILTLIFFLAKDLISLKKAFSII
IGSRILGVLNQIFVKLDLIEIKYVNFYRQDGFILRSDLGLGHPNFTHNFALTIFYL
IVLVYKRLKPVMLVFLTNLYLYQVTFSTGYIIVTLFVLYVTNKSLIKRVFMKL
APVQFPLLVFTLSSTIPFNSNFVOKLDVLLNGRHLVAHLQVLDGIGTFRGNSFKETS
VFDNSYMLSMYGVVLMFCMIIYIYKSKIIILIELQLLFIIMSILITFESFPYSPV
VMNISLWVGKIFCDGIEPIKKEFTIVNNI"
complement (8147. .8638)
/gene="cpsG"
/note="synonym: SAG1169"
complement (8147. .8638)
/gene="cpsG"
/note="similar to GP:5823214; identified by sequence
similarity; putative"
/codon_start=1
```


Matches 3323; Conservative 0; Mismatches 3382; Indels 185; Gaps 31;

| | | | |
|----|-------|------------------------------------------------------------------|-------|
| Qy | 16 | TGGCATTATTTGATATGATAGACGTTTGCAAATTTCTGCAATCTTAACAAGTCATATAACCAA | 75 |
| Db | 10465 | TAGCGATGAATCAAAACAGTTGTGGTTTTATTTTTCTGCAAGTTTGACATTAACATTAATTA | 10406 |
| Qy | 76 | ATGCTGATTTAAATCG---TTCTGGGAATTTTTATCATATATGATGGTTCAATATTTTGCAT | 132 |
| Db | 10405 | CTCCCAATTTTAAAAGCAATAAAGAATTAATGTGTTGTTCTATTTGATACATATATTTGTCT | 10346 |
| Qy | 133 | TTTTTATATCTCGTATGCCAGTTTGAATTTTGATATAGAGCTAATCTGATAGAGTTTGGAAA | 192 |
| Db | 10345 | TTTTATCTTTCGATTTTTACAGAGACTTTTGGANGTCGGCTATCTTGAAGAGTTTAAAA | 10286 |
| Qy | 193 | AACATTTAACTATAGTATAATATTTGCAATTTTTCTTAGCGCAGTATCATTTTTTGTGG | 252 |
| Db | 10285 | TGGTATTTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCAATTTTTTTATTTTTTA | 10226 |
| Qy | 253 | AGAATAATTTGCGACITTTCAAGACGTGGTCCGCTGATTTTCATATTAATAAACITTCGTTT | 312 |
| Db | 10225 | AAAACTCTTTTACAACGACACGACTTTCCCTTTTTTACTTTTTATGCTATGAATTCGATTT | 10166 |
| Qy | 313 | TGTTATACCTATTTTAAGCTAAATTTAATAGCAGTTTAAAGGATAGCTTCTATTTTTCGACAA | 372 |
| Db | 10165 | TATTATATCTATTGGAATTTCAATTTTTTAAATATATTACGAAAAATTTCTTAGCGTAAGTTTT | 10106 |
| Qy | 373 | TCTATCAAAAAAGACGATTTCTAATTAACAACGGCTGAACGATCGGAAAAATATGCAAGTTT | 432 |
| Db | 10105 | CACGAGTACCAGATTTGTTTTCATACGNAATAGGATTTCTTTATCAAAAAATGACCTTTA | 10046 |
| Qy | 433 | TATTTGAAATCACATAAACAAAATTTCAAAAAAATCTTGTTGCAATGGTAGTTTGTAGTACAG | 492 |
| Db | 10045 | GGAATAAAT---ACGACCATAATATATCGCTGCTGTAATCTTGGATTCCTCTGAAAAAG | 9989 |
| Qy | 493 | ABATAGATAAAATTAATTTATCATTAACCGCTCTATTATTTCTGTGGAAGAGCTATAGAGT | 552 |
| Db | 9988 | ATTGTTATGATTTTGAACACATACTCGTTAAGGANAATAAACAAAGATGCTCTTACTTCAG | 9929 |
| Qy | 553 | TTTTCAACAGGGGAAGTGGTCGACACAGTCTTTTATAAATCTACCAAGTCGAGTTTTTAGACG | 612 |
| Db | 9928 | AGTTAACCTGCTTAACTGTTGATCAAGCTTTTATTAAACATACCCATTGNAATTTTGGTA | 9869 |
| Qy | 613 | TAAAGCAATTC-----GTTTTCAGATTTTGAGTTGTTAGTATTTGAATGAAGCGTTG | 663 |
| Db | 9868 | AATACCAAAATACAAGATATTTAATAGCATTTGAAGCAATGGAGTGATGTCAATGTTA | 9809 |
| Qy | 664 | ATATTAAATTCATTCGGTTTTACTGCGTTGMAAAACAAAAAATCCNACTGCTAGTGTACC | 723 |
| Db | 9808 | ATGTAGAGGCATTTAGCTTTGATAATATAGGAGAAAGCGAATCCAAACTTTTGAAGGAT | 9749 |
| Qy | 724 | ATAGCATGTGAACTTTTTCCACAAATTTTTATAAGCTAGTCAATCATGATGAACACGAC | 783 |
| Db | 9748 | ATAGTGTATTACATATTTCTATGAAATTTCTATAAATATAGTCACCTTATGCAAAACGAT | 9689 |
| Qy | 784 | TTTTGGATAPACTCGGACGGTATGTCGGGTAAATTTATTTGTGGTATAGTTCTTATTTGT | 843 |
| Db | 9688 | TTTTTGGATATCACGGGTGCTATTATAGTTTGTCTCATATGTGGCATTTGTGGCAAATTTTTC | 9629 |
| Qy | 844 | TAGTTCCAATTTTCGTAGAGATGGTGGACCGGCTATTTTTTGTCTGAGAAACGAGTTTGAC | 903 |
| Db | 9628 | TAGTTCCACAAATCAGAAAAAGATGGTGGACCGGCTATCTTTTCTCAAAATAGAGTAGTGC | 9569 |
| Qy | 904 | AGAAATGACGCATATTTACATTTCTACAAGTTTTCGATCGATGATATGTTGATGCTGAGGAC | 963 |
| Db | 9568 | GTAATGTGTAGGATTTTATAGTATCTATAAATTCAGATCAATGCGAGTAGATGACAGAACAA | 9509 |
| Qy | 964 | GCAAAAAGACTTTGCTCAGCCAAAAACAGATGCAAGGGTGGGTATGTTTTTAAATATGGGAA | 1023 |
| Db | 9508 | TTAAGAAAGATTTATTAGTTTCACAATCAATGACAGGG----CTAAGTTTAAAGTTAGA | 9454 |
| Qy | 1024 | AAACGATCTTAGAATTAATCTCAATTCGACATTTTCATACGCAAAAAACAGTTTGTAGACGAGT | 1083 |
| Db | 9453 | CGATGATCTCTAGAAATTAATAAATAGAAAAATTTATTCG-AAAAACAAGCATAGATGAGT | 9395 |

Db 6234 TTATCCTATC-TGTAATGATCTCTATGTAATCAGACAAATAAAATTTTATATGTTGAT 6176
Qy 4356 GAAGAACGATGGAAGAAATTAATAATATTACAAGAAATTAATGCGGTCTAGCAGATGCTCG 4415
Db 6175 TGTGATGATGTTCTGAGGACAGTACTAAGGAATTCGTAATTAATAAAGAGAA 6116
Qy 4416 AAATTTCCGACTAGAACATGCAACAGGTAAATATATGCTTTTGTGCGATCTGATGACTA 4475
Db 6115 TAAAGTTAGTATGTTATATCTGTGTA---TAAACGTAATGCGGTAGCAATTCAGCCATATA 6059
Qy 4476 TATGAAGTTCGAATGTTCCGAGAGATGCATGATAATAAATGAGTATATGCGGATAT 4535
Db 6058 TCTGCAATGAGATATATGCAACCAAGTGAATATCATGTAATGTTAGATAGTATGATG 5999
Qy 4536 AGCAGAGATAGATTTTGTGTTAGTAGACGAAACCGGTATACAAGAGAAAGAAATAG 4595
Db 5998 GTTATTAGAGATGCAAGTTGAATTTATTTTAAAGATTTAGAGACCTAAAGTTGCTTAA 5939
Qy 4596 TAAATTTTCATGCTTTAAACGAGAGAGACCTGTAAAGAAATTTTGTGAGGATCTAAATAT 4655
Db 5938 TAGATATGAGCCCT-----TGTTTATCCTCGATATTCATTAACCAAGGTAAATAATG 5885
Qy 4656 AGAAATATATGTTGGTCCAGCTTTTATTCAGGATATTAATAAGATATAAATTCCTCA 4715
Db 5884 GTTGAATCCCAAAATATTAGAGTTTAAATTTTAAAGATTTAAATATATCACTTAA 5825
Qy 4716 AATTAAATATAGAGTATGTTGAGGATTTGCTTTTAAATTTGAGGCTTTGAACAAATG 4775
Db 5824 AATTGAACCTTGTATGTTATTAATAA--TGCCTATTTAGTAGATTTTGAATTTCTTGT 5767
Qy 4776 AACAGCTGATGATGATATCAGAGAAATAATTAATTAATATGATCTTCGTAAACAGTTC 4835
Db 5766 TTCGAAGGGAGAAATTTCTTATCAGAGAAATAATGATATATATTTATCAAAAGGGA 5707
Qy 4836 GCTTATTAATCAGAAATTTCTATTAATAATATGATTTAGTCAAGATGAGAAATTA 4895
Db 5706 TACTTTTGTCCCAAAATAGGAAATCTATTG---TTTGTACTTCTTAGAAGATGGTTT 5651
Qy 4896 CCCCTTTAAGTTAAAGAGAGTTPAGTCATTTATTTTCGATGCAAAAGTTATTAAGAGAA 4955
Db 5650 GACTTCTAATATTTTAAATTTATGAGAGAAATAATTTCAAGGACTATTTTTCATTAAGA 5591
Qy 4956 GGTAAATGTTTAAACAAATGTTATTCACAGATGTTTGGATATGATGTTCTTGCCTAAT 5015
Db 5590 AAATTCATATATGATGTAATGCTTTTCTTCAATATATTTGATCGCTGCTGCTCAGC--T 5533
Qy 5016 ATTAGAGTCTTATCGAAAGAAATACGATAGATATCCATTTTAAAGCGAAAGAAATATTT 5075
Db 5532 ATAAAGATAAAATGAAACATACAGGCTTTGAAATGACTACTTTTAGGGGTAAATACCTACA 5473
Qy 5076 ATCAAGAAAGCATTTAGTTCGTTGTAATTTGATGAAATTTTCGCTTAACTATATGTAAT 5135
Db 5472 CTAAG-----AGTGAAGAGCTCGGTGGGAAATTTTGTGGGTTTGTAGCTATCTT 5422
Qy 5136 GTTATATAGAAATTTCAAAAGCAGTAGAGGTAAAAATGGATATAATTTAGTGTATTGTT 5195
Db 5421 TGGAGGTAGCAAGTTTAAAGAGCGAGTAAGAAATGGATTAATTTCAATAATAGTG 5362
Qy 5196 CCAAGTTTAAATGATAGATAATTTAATAGTATGTTGATAGAAAGCATTTAATCAAAAT 5255
Db 5361 CCGTATATAATGGAGAAATATATTTGGGAGATGCTAGATAGTATTTCTTTGAGCAAACT 5302
Qy 5256 TATAAATATAGAAATATTTATGATAGATGATGCTCTGATAGATGATTTCTGCTAAATA 5315
Db 5301 TATCAGAAATTTAGAGATTTATCATAAATTTGATGATGCTCTAGTGAATCGCAGGGGATAT 5242
Qy 5316 TGCAAGGAATATGCAAAAGAAAGATAAAGAGTAAAAATTTTTTCACTAATCATAGTGA 5375
Db 5241 TCGGAAAGATATTTTGGAGAGACAGGCGAATAAATAATTTCTATCAAGAAACAGAGC 5182
Qy 5376 GTATCAAAATGCTAGAAATCATGGAATAAAGCGGAGTACAGCTGGAATATATTTGTTT 5435
Db 5181 CAATCAGTTGCAACGTAACACCGAGTACTTAGGTGATCAGGAGATTTGGATTTGCTTTT 5122

Qy 5436 GACTCTGATGATGTTGTTGATAGTAGATAGTAGAGAAAAATTTATATTTTAAATATTATAAAA 5495
Db 5121 GACAGTGATGATGTTTATCTCTTCAATTCAGGTTATGTAACAACATACAGAAAGCT 5062
Qy 5496 AGTAGAGTATGATTTATCTGGTGTGTTGACCTACTTTTTCAGAAAAATATAAATAATTTT 5555
Db 5061 ACCAATCCGATATGTTTAACTAGTATGTTAACTTCAACAATCTTCAACTACTAGT 5002
Qy 5556 GAAGTGAATATCCAAATATTGATTTTGAAGCAATTAATACCGTCGACGACATGCGAGAA 5615
Db 5001 ATTAATTCACAGTATTTAAAGAGATTAACCTTTATCTTTAGAGGTGCGTTTGAAGAA 4942
Qy 5616 AAAAAATTTATGAATTTGTATATAAATAATTTTCTACTCTCTCTTTGTAAACTATAT 5675
Db 4941 ATGTATTATGGAACAACTTATGGAGTTTCCCGCTTAGCAA-----AATATATCCGAGAA 4887
Qy 5676 AAGAAAGATACATAACAGATCTTTTTCAGAGAAATCAATCGTTAGGAGAAAGATTTACTT 5735
Db 4886 GTAAATTTATGAGTAATCCATATCCAGAAAGGAAAAATTCATGAAGATATGGATACAACCT 4827
Qy 5736 TTTAATCTGCAATTTTAAAGAAATATAGATAGATGTTAGTATTTTACCTGAACATCTTAT 5795
Db 4826 TTAAGCTTATATCTTGGCTTCAAAAATAGCAGTTTGTGATTTGTCTACTCGAGTGTGT 4767
Qy 5796 TTTTATAGGAGAGGTATACCTAAGTACAGTAAATTTCTTTTAAAGAGAGGTGTGTTTTGCAA 5855
Db 4766 ATTTTAGCGATAATAGTACTACTAGGACAAATTTCAATGAGAGAAATGT-TGTATTTTTT 4708
Qy 5856 TTGGAATTTTGGAAAAACAAAGTATGATGTTTGTAAAGAAATATATGTTGAGGATTTT 5915
Db 4707 GAGGCAATTCAAAAATAATTTGTATTTTATAAATCTAAATTTTCCACATAATACATCTTTA 4648
Qy 5916 GACGTATCAATTTGTTAAAGATACATATCGTTGGCAAGTATTTTATATAGCTTACTAATG 5975
Db 4647 ATCTCTGCTGTAATATATATGAAAGTATTTGTTGGAAATAGATATATGTTGGAGATGAT 4588
Qy 5976 TTTAAATACGGAACACAGTCTATTTTGGACAAATTTTAAATTTTGTAGAAATCTTTTATAA 6035
Db 4587 GATTTTAAATATATGATACAGTGTGATTTATACAGAAAAAATATAGAAATATTTTAAG 4528
Qy 6036 AAATATATTTTAACTGTTTAAAGATCTAACAAAAATCTTTGTCTTAAAAATTTTGT 6095
Db 4527 ACTATCTCTTAAATAATAGGATTAGTGTGAAGGAGAGGTTAAATAATATATTTATTTAT 4468
Qy 6096 ATAGAAATTTGTTTCGAACAAAGTTTAAAGAAATATATGTTTATATAGAGATATC 6155
Db 4467 TCCTCTATAAGATATTTCACTATTTGTAAAGAAAAATA-----TATAATTTGAGGTTATC 4415
Qy 6156 ATGATACTATTTAGTAAAAATTTCTATATTTGTAATTTGTAATATATATGTAAGAAAAATTTTA 6215
Db 4414 AAGAA-----ATGGATTAAGTTAGTATCTATTTCCAGTTTATACGTTCAATCTTCTCTA 4359
Qy 6216 TCTAAATGATATAGATAGCATTTGTAATACAGACCTACAAACATATAGAGATTTCTTCTG 6275
Db 4358 AACGAGTATAGAAAGTGTTTTAGCTCAAACTTATTTCAAAATTTAGAAATTTTCTAGTT 4299
Qy 6276 AATGACGATAGTACGATTAATTTCCGAAGAAATTTGTTTGTAGCATATGCGAGAAAGATAGT 6335
Db 4298 AACGATGCTCTACAGATAATTTCCGGGGATATCTGTGATTTATTTTCAGAAATATAGATGG 4239
Qy 6336 CGCAATTCGTTATTTTAAAAAGAGAACCGCGGCTATCCAGATGCCGTAAATTTATGCGCAT 6395
Db 4238 CGTATATTCGTTTTCACAAAAATAATGTTGTTTATCAGATGCTGCTGTAATTTATGGAAT 4179
Qy 6396 AGTCGCGCAAGGTGACTACTTAGCTTTTATAGACTCAGATGATTTTATTTTCAATTCGGAG 6455
Db 4178 AGTAGAGCTACTGGGATTTATTTTACTTCTGATTCGGATGATTTCTATATAAAGAA 4119
Qy 6456 TT---CATCCAACTTTACAGAAAGCAATTCAGAGAGAGAAATGCCCTTGTGCGAGTTGCT 6512
Db 4118 GATCAATCCGAGAGATGTTTGAATTTTAGCCGAGAAATAAATTCAGAAATTTGTTTAGGG 4059


```
Qy 6513 GGTATG-----ATAGGTAGATGCTCGGGGCAATTTCTTAACAGCAGAGCCGCTTCCT 6566
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4058 TGTATGTTGAAAAGAGAACACATATTATAATATCGTTTATAGAGATGAATGATA 3999
Qy 6567 ACAATCAGGCTGTTCTGAGCGGAGGAATGTTTGTAAAAGCTGCTAGAGCGGATGGT 6626
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3998 GAAACTATTAGTCCAGTACAGCAATACAAATATTTTATAA---TTATGATGCTATAGA 3942
Qy 6627 CATCGCTTGTGGTGGCTGTATAAATCTCTATAAAGAGACTATTTTGAAGATTTTCGA 6686
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3941 GCTATTTTAACTGTAGCACATAATAAATATATAAGAGAAATTTATTTCTACGCTGTC 3882
Qy 6687 TTTGAAAAGGTAAAGATTCATGAAGATGAATCTTCACTTATCGCTTGCTCTATAGATTA 6746
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3881 TATCCAGTAGGAAGTTACAGAGATGAATTTTAACTATAAATTTATTTTGAAGGCC 3822
Qy 6747 GAAAAGTTGCAATAGTTAAGGATGCTTGTAATTTATGTTGACCGAGAAATAGTATTC 6806
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3821 AAAAATATAATTTCTTTTAGATAAATACATTATGCATATCGTATTAGAGAAAATAGCATA 3762
Qy 6807 ACAACTTCTAGCATGACTGACCATCGCTTCCATTTGCCCTACTGGAAATTTCA 6856
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3761 ATGACTGGTCTCTACAATATTTAAAAGGCTACATGCTGTAGAAGCACTTAA 3712

RESULT 10
LOCUS CQ655070 349980 bp DNA linear PAT 02-FEB-2004
DEFINITION Sequence 12027 from Patent WO234771.
ACCESSION CQ655070
VERSION CQ655070.1 GI:41687947
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1
AUTHORS Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 12027 02-MAY-2002;
Chiron S.p.A. (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
    Location/Qualifiers
        1..349980
            /organism="Streptococcus agalactiae"
            /mol_type="unassigned DNA"
            /db_xref="taxon:1311"
        /note="seq 10967 to big: splitted in-seq 10.967, from
0.000.001 to 0.349.980-seq 12.025, from 0.300.001 to
0.649.980-seq 12.026, from 0.600.001 to 0.949.980-seq
12.027, from 0.900.001 to 1.249.980-seq 12.028, from
1.200.001 to 1.549.980-seq 12.029, from 1.500.001 to
1.849.980-seq 12.030, from 1.800.001 to 2.149.980-seq
12.031, from 2.100.001 to 2.160.266"

ORIGIN
Query Match 11.4%; Score 798.8; DB 6; Length 349980;
Best Local Similarity 48.2%; Pred. No. 9.7e-82;
Matches 3323; Conservative 0; Mismatches 3382; Indels 185; Gaps 31;

Qy 16 TGGCAATATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAAACAAGTCATATACCAA 75
Db TAGCGATGATTCAAACAGTTGTGGTTTATTTTCTGCAAGTTGACATTAACATTAATTA 247176
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATAATGATGGTTTCATTTTTCAT 132
Db CTCCCAATTTTAAAGCAATAAAGATTATTGTTTGTCTTCTATGATACATTTATTTGCT 247116
Qy 133 TTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAACTGATAGAGTTGAAA 192
Db TTTATCTTCTGATTTTTTACAGAGACTTTTGGAGTCTGGGCTATCTTTGAAGAGTTTAAA 247056
Qy 193 AAACATTTAACTAGTATATAATATTGCAATTTTCTTACGGCAGTATCATTTTGTGTG 252
```

```
Db 247055 TGGTATTTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTATTTTATTTT 246996
Qy 253 AGAATTAATTTTCGACACTTTCAAGACGTGGTGGCGGTATTTTCATTAATAAATCACTCGTTT 312
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246995 AAAACTCTTTTACACGACACGACTTTTCCCTTTTTTATTTTATTTGCTATGAATTCGATTT 246936
Qy 313 TGGTATACCTTATTAAAGTAAATTTAAAGCAGTTTAAAGGATAGCTTTCTATTTTCGACAA 372
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246935 TATTATATCTATTGAAATTCATTTTAAAATATTATCGAAATATTCTTACGCTAAGTTTT 246876
Qy 373 TCTATCAAAAAAGAGCAATTTCTAATTTACACGGCTGAACGATGGGAAAATATGCAAGTTT 432
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246875 CACGAGATACCAAAGTTGTTTGTATAACGAATAAGGATTTCTTTATCAAAATAGACCTTTA 246816
Qy 433 TATTTGAATCATAAACAATAATTTCAAAAAATCTTGTTCGATTTGGTAGTTTATAGGTACAG 492
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246815 GGAATAAAT---ACGACCATAAATTATATCGCTCTGTATCTTTGGATTCCTCTGAAAAGG 246759
Qy 493 AAATAGATAAAAATTAATTTTATCATTTACCGCTCTATTATTCTGTGGGAAGACTATAGAT 552
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246758 ATTGTTATGATTTGAAACATAAATCGTTAAGGATATAAACAAGATGCTTCTACTTCAG 246699
Qy 553 TTTCAACAAGGGAGTGGTCGACCACTCTTTATAAATCTACCAAGTGAGTTTTTAGACG 612
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246698 AGTTAACTGCTTAACTGTTGATCAAGCTTTTATTAACATACCACTTGAATTTTGGTA 246639
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTCGAGTTGTTAGGTATTTGATGTAAGCGTTG 663
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246638 AATACCAAAATACAGATATTATTAATGACATTTTGAAGCAATGGGAGTGAATTTGCAATGTA 246579
Qy 664 ATATTAAATCATTTCCGTTTTTACTGCGTTGAAAACAATAAATAAATCCCACTGCTAGTGCAC 723
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246578 ATGTAGAGGCATTTAGCTTTGATAAATATAGGAGAAAGCAATCCCAACTTTTGAAGAT 246519
Qy 724 ATAGCATTTGTAATTTTTCACAAAATTTTATAAGCCCTAGTCATATCATGATGAACGAC 783
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246518 ATAGTGTATTACATATCTATGAAATCTATAAATATAGTACCCTTATAGCAAAACGAT 246459
Qy 784 TTTTGGATATACCTCGGAGCGGTAGTCGGGTTAAATATTATTTGGTATAGTTTCTATTTTGT 843
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246458 TTTTGGATATACCGGGTGTCTATTATAGTTTGTCTCATATGTGGCATTTGTGCAATTTTC 246399
Qy 844 TAGTTCCAATTTTCTGTAGAGATGGTGACCGGTATTTTGTCTCAGAAACGAGTTCGAC 903
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246398 TAGTTCCAAAATCAGAAAAGATGGTGACCGGTATCTTTTCTCAAAATAGAGTAGGTC 246339
Qy 904 AGAATGACGCATATTATTACATTTCTCAAGTTTTCGATCGATGTATGTTGATGCTGAGGAGC 963
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246338 GTAATGTTAGATTTTGTAGATTTCTATAAATTCAGATCAATGCGAGTAGATGCAACAAA 246279
Qy 964 GCAAAAAGACTTGTCTCAGCCAAAACAGATGCAAGGTGGGTATGTTTTAAAATGGGAA 1023
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246278 TTAAGAAAGATTTTATTAGTTTCACAATCAATGACAGGG-----CTAATGTTAAGTTAGA 246224
Qy 1024 AAACGATCTAGAAATTAATTCGAATTTGACATTTTCATACGCAAAAACAAGTTTACAGAGT 1083
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246223 CGATGATCCTAGAAATTAATAAATAGGAAATTTATTTCG-AAAAACAGCATAGATAGT 246165
Qy 1084 TACCACAGTTTTTAAATGTTTTTAAATTTGGCGATATGAGTCTAGTTGGTACAGCTCCACTA 1143
Db TGCTCAATTTCTATAATGTTTTTAAAGGTGATATGAGTTTAGTAGGAACACGCCCTCCCA 246105
Qy 1144 CAGTTGATGAATTTGAAAAATATCTCTGGTCAAAAGACAGATGAGTTTAAACAG 1203
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246104 CAGTTGATGAATATGAAAAGTAAATTTCAACGCAAGACGACGCTTTAGTTTAAAGCAG 246045
Qy 1204 GGATTCAGTCTCTGCGAGTTAGTGGTCTAGTATATATACAGACTTCGACGACGCTAG 1263
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246044 GAATCACTGGTTTGTGGCAAAATCTGTTGAGAAATAATATTACTGATTTTGTGAAATCG 245985
Qy 1264 TTCGTTTGGACTTAGCATACATTTGATAAATTTGGACTATCTCGTCAGATATTTAAATTTTAT 1323
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


Db 245984 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTAAGATTATTC 245925
Qy 1324 TAAAGACAGTGAAGTCTGTTATTTGGTGAAGAGGGAAGTAAGTAAAGATAT 1373
Db 245924 TCCTAAACATTAAGGTAGTTTATCTCGGGAAGAGGCTAAGTAAAGGTAAAGGTTTGAAG 245865
Qy 1374 -----ATGAAAGTTGTTGGTGGTCTTTCAGGGGGAATTTGACTCACTTGTATTT 1426
Db 245864 GAATATATGAATAATTTCTGTTGGTGGTCAAGTGGTGGTCACTAGCACACTTTGAACCT 245805
Qy 1427 GTTAAACCGTTTGGGAAGGAAGAGACGTTTTTGGGTAAACATTTGATATAAGAGGATGC 1486
Db 245804 TTTGAAACCCATTTGGGAAAGAGAGATAGTTTTTGGGTAAACCTTTGATAAGAGATGC 245745
Qy 1487 AAGAGTCTTTTGAAGAAATGAATAATGTATCCATGTTACTTTTCCAAACAAATCGCAATCT 1546
Db 245744 TAGGATATCTAAGAGAGAGATGTATATCAATTCATTTCTTTCCAAACAAACCGTAATGT 245685
Qy 1547 CATTAATTTAGTGAATAATATCTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606
Db 245684 CAAAAACCTGGTAAAAAATACTATTCTAGCTTTTAAAGTCTTTAGAAAGAAAGAACCCAGA 245625
Qy 1607 TGTTATATTTTCACTGGTGGCGGCTTGTCTGCCCTTCTTTTACATCGGAAACATTT 1666
Db 245624 TGTATCATATCATCTGGTGGCGCTGTAGCAGTACCATTTCTTTATATTTGGTAAGTTATTT 245565
Qy 1667 TGGAGCAAGACGATTATATTTGAAGTATTTGATCGAGTTAATAATCTACATTAACCTGG 1726
Db 245564 TGGTTGAAGACCGTTTACATAGAGTTTTCGACAGGATGATTAACCACTTTTGCACAGG 245505
Qy 1727 AAAAAGTATTTCCCGTAAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGGT 1786
Db 245504 AAAATTAGTGTATCTGTTAAACAGATAAATTTATTTGTTTCAGTGGGAAGAAATGAAGAAGGT 245445
Qy 1787 ATATCCTAAATCTATTAATCTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAT 1846
Db 245444 TTATCCTAAGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTACACAGTGGGACA 245385
Qy 1847 CATGAACAAACAGTTTAAATCGATTGATAAAGAGATTTGATTTATGAAAAAATGGAAGT 1906
Db 245384 CATGAACAGCAGTTCCACCGCTTATTAAAGAGTTGATGATTAAGAGGACAGGTGCT 245325
Qy 1907 ATAACCGACGAATAATTTATTCAAACAGGATATTTCTGACTATTTTCCAGAAATTTGCAAG 1966
Db 245324 ATTGATCAAGAAGTGTCTCATTCAAACGGTGTACTCAGACTTTGAACTTCAAGATTTGTCAG 245265
Qy 1967 TATAAAAAATTTCTCAGTTTACAAAGAAATGGAACAAATATTAACAAATCAGAGTAGTT 2026
Db 245264 TGGTCAAAATTTCTCTCATATGATGATATGAACCTTTACATGAAGAAGCTGAGATTGTT 245205
Qy 2027 ATTTGCCAGGAGGCCCGCTACTTTTATGAATCAATTTATCCAAAGGAAGAAACAAATTA 2086
Db 245204 ATCAACATGGCGGCCCGAGCGTTTATGAATGAGTTTCTTAAAGGAAAGAAACATAT 245145
Qy 2087 TTGTTTCTAGACAAAAAAGTATGGTGAACATGTAATGATCATCAAGTAGAGTTTGTGA 2146
Db 245144 GTGGTCTCAGACAGACAGTTTGGAGAGCATGTGAATATCATCAGTGGACTTTGTT 245085
Qy 2147 AGAAGAAATTTTCAAGATAATAATATTTTATATAGAAATATAGATGATTTGTTGAA 2206
Db 245084 AATAAGGTAAAAACAATGTATAATTTTGTATATCGTTGTAGATATTTGAAGGTTTCAAAAT 245025
Qy 2207 AAAATATTTGAAGTTTCTTAAGCAAACTAACTTTTACATCAATTAATTAATTTT 2263
Db 245024 GTAGTCTATAGGGAATGATGAATCGTCCGTTTTTGAAGAACTAAATAGTAGTAATTTTAT 244965
Qy 2264 GAAAGATTTAAACAAATAGTTTGAAGAAATTTAATGAGGATCAAGAAAAATGAATAATAAAA 2323
Db 244964 GAAGAAATTTAAGGTAATATTAAGAGATTTGCGATGATAAATCAATATAAAACTCTTTAT 244905
Qy 2324 AGATGCATATTTGATAATGGCTTATCATAAATTTTCTCAGATTTTCTGAGAGGGGATAC 2383
Db 244904 TTATATGCAATATTTTATAGTTAAATTTTAAATCACTGGGTTTTAGCGGAGGAACTC 244845

Qy 2384 AGATATTTATCATCTTCTCTCAGGAGAAATGCACACCAATTAGTTCTCTCAGAAATACCTGTAT 2443
Db 244844 AGCTTACAAAATAGTG-----ATGTTAGTTGCAA 244816
Qy 2444 AATTATTTTAAATATTTCTCAGGATTTATATGTTGAATTTTACAAAAGATGACGCAAAATAT 2503
Db 244815 TTTTACTGTGTGGAATAAATTTTTTATTTAGATAGCCTTTTATTTTGAAGAAGAAACATCG 244756
Qy 2504 AAAGAAATAGGATATATGAACGAGTTTAAATGTTTACAGATTTATTCCTAATATATCAGAA 2563
Db 244755 TGATCATCTTTTATTTATTTATCGGACCAATTTTGAATTTATCT----- 244711
Qy 2564 AAAAATTTGATAATGATCTGTTTGAATTTTATTAAGAATGTATCGAGCTTTTGAATAC 2623
Db 244710 ----TTGTTTCATAAGTTACTTTTATATAATACTTTTATTTTCTAGCATTAAGGAC 244655
Qy 2624 TATTTACAAAGATTTGTTTATTTGATAGATAAATAAACAATGCTTAAGAAATTAAGATTTG 2683
Db 244654 ATCTCTCAAAAAAGCTTTCTCTATAATAATAGGATCGCGTATTTTGGGAGTTCTATTA 244595
Qy 2684 GTTCTAATTTGGTTCCTCCACATGATTTTGTGGCAATTTCTTTATCAAAATGAAAAAG 2743
Db 244594 AATCAAAATTTTGTGAATTAGATTTAATAGAAATTAAGTATGTCATTTTATAGGAT 244535
Qy 2744 AAACAGCTTATTTTAAAGTAATCTAAATGCTCAGATGAATATTTATATACAGACAATTA 2803
Db 244534 GGACAAATTTATCTCGAGAGTGA-CTTAGGTTTGGTCACTTAATTTATTCATTAATTT 244476
Qy 2804 TAGAAAAATATGAATTTTCAANTAGATATCTAAATATGAAATTTGAA-----ATTTAGATATATAA 2860
Db 244475 TTTTGCTCTACTATTTTCTTGATATTTGTACTCAATTTATAACGACTTAAACCTTTTGT 244416
Qy 2861 AGTGGAAAAATCAACATCTTCTCTATTTGCTTTTACAGATGATTTCTATTCATGAATTC 2920
Db 244415 GATGGTTTATTTTAAATTAATTTATTTGACCAATATTTATTTGACCAATATCTTTTCAAGCAAG 244356
Qy 2921 TAAATGCAAGAAATTTTAGTTTTTATTTGCTAGAAAGTTAAAAATAGAAATTAATCTA 2980
Db 244355 GTATTATATCGTAATTTTATTTATTTGTTACTCATTTATGTCGCAAGAAATAGCTTAATAA 244296
Qy 2981 AA----TTTAAAGAAATTTTACTTAAAAAAATAAATAGTTGATTTTGTGAGAGTAATGTA 3036
Db 244295 AAGAGTATTTATGAATTTAGCACCTATGTACAAATTTTTTTTTTATTTAGTATTTTACCTTTT 244236
Qy 3037 TGTTTAAATTTTAAATATGACCCGGAATATTTTATTTTAAAGTACTCTCTGGTTGATTA 3096
Db 244235 GAGTTCTCAAAATTTTTTAAATTTTAAATTTTGTTCAAAAATTTAGATTTCTTTTAAACAG 244176
Qy 3097 TTTTATTTCCAGAGCAAAAGTATGTTATTTTATTTTAAATTTTATGAAATTTTATTTTCT 3156
Db 244175 TAGATTACATCTGCTCATTTTACACTTTGTAGATGGTTTAACTCTTTTGGAAATAGTTT 244116
Qy 3157 ATATPAAATTTTGAACCTAAGCTAAATTTTAAAAAAATGAATTTTATTTTATTTAT 3216
Db 244115 TAAGAAACAGTGTCTTATTTGATAATAGCTACTCTATGTTATTTAGTATGATGTTGTT 244056
Qy 3217 GGTCTATATTTATGTTTGTGTTTTCAGTAGTCACAGTATGTTTGTGGAATAAATTTTGAA 3276
Db 244055 AGTACTTACCATTGTTTGTATGATACTATTTATCTATATAGTAAAAAG-----TAA 244003
Qy 3277 GATTATTTGACAGTTTCTCTCTCCCAATTTTGGATTTTGGCAATGATTTATTTATTAAT 3336
Db 244002 TCATAATTTGAACCTTCAACTACTCTCTATTTTAAATGTTCTATTAATTTATTTACTGAAAGTT 243943
Qy 3337 TGTATTTCAATTTTAAATTTGATTTATAAAAAATTTAAAAAATAGTATCTTTTATTTT 3396
Db 243942 TTTATCCAGTGTGTATGAATATTTAGTTGCTAGTTTGTGGTAAATAATTTTCTGTATG 243883
Qy 3397 TAGTTTTATTTAGGTATATCTGCTGATTTATATTTTCAAAATTTGGGAAAGATTTTGTATTT 3456
Db 243882 GTATCGAACCTATAAAAAAGGAATTTTACTATTTGTGAATAAT--ATATGACATTTTGTCT 243825

| | | | | | | | |
|----|--------|------------------------------------------------------------------|--------|----|--------|----------------------------------------------------------------|--------|
| Qy | 3457 | TAGACAGACACCTTTATAGGACTAGACTATCTTATAACAGGGCTCAAAACAAAGTTGGTTG | 3516 | Db | 242768 | GTTATTAGAGATGCGAGTTGAAATATTTTTAAAGATTTAGAGAGCCCTAACGTTGCTCTAA | 242709 |
| Db | 243824 | TGATATGGCAGGAGTTAAGNAGGAAATGATACCTTAAAGTTATACATTTATGTTGGTTT | 243765 | Qy | 4596 | TAAATTTTCATGCTTTAAACGAGAGAAGAGACTGTAAAAGAAATTTTGTGAGGATCTAATAT | 4655 |
| Qy | 3517 | GCTTTATGAACTATCCCTACGTTAAATACCACTACAATTTAGTTTTCAATTCGCTTAATCT | 3576 | Db | 242708 | TAGATATGTAGGCT-----TGTTTATCCTCGATATTCATTTAAACCAAGGTAATAATTG | 245655 |
| Db | 243764 | GGAGGAAATCCCTTACCAGATAATTTAAAGAAATATATAAAACCTTGAGAGAACAAATGT | 243705 | Qy | 4656 | AGAAATTAATGTTTGGTGCAGCTTTATTCACAGAGATTTATTAAGAGATATAAAATTTCCA | 4715 |
| Qy | 3577 | TTGCACCTATTAATAAATAAATAATGCAACAATTTTTTTTCTTGTGTCCTGCTTTTATACCGA | 3636 | Db | 242654 | GTTGAATCCCAAAATATTGAAGTTAATATTCCTGATTTAAAAATATAAAATATCACTTAAA | 242595 |
| Db | 243704 | CCGATTTATGAAATATTAGTAAGAAATGAGCATAAATTTATGATG----- | 243661 | Qy | 4716 | AATTAATAATAGAAATTTGGTCAGGATTTGCTTTTTTAAATTTGAGGCTCTTGAACAATGT | 4775 |
| Qy | 3637 | TCATATTAAGTGGATCCGAAATTCGTAGTTTATCGCTAGCAATTAATTAATATATGCTTGT | 3696 | Db | 242594 | AAATGAACTTGTTATTTATTAATAA--TGCTTATTTAGTAGATTTTCGAATTTTCTTGT | 242537 |
| Db | 243660 | --TAGTAAATGTTTTATGAGAGAAGCATATACCTAAGAGAAATTTTGCCTTATGTTTCT | 243603 | Qy | 4776 | AACACGTGTAGTTAGTTGATACCTAGAGAAATTAATATAATATGTCAATTCGTACAGTTC | 4835 |
| Qy | 3697 | TATGGAGATATATAGG--TGGAATAATTTGCTTTGGATATAAAAGCTAAATAGTAATATTTGTA | 3755 | Db | 242536 | TTCGAAGGGGAGAAATTTCTTATCAGAAGAAATAATGTATATATATTTATCAAAAAGGGA | 242477 |
| Db | 243602 | GACTATGCAAGATTTGGATATTTATTTATACTTATCGGGGTTCTATCTAGATACATGATG | 243543 | Qy | 4836 | GCTTATTAATCAGAAATTTCTCTATAAATAATATTTAGTTAGTCACAAAGATTTGGGAATTA | 4895 |
| Qy | 3756 | ATACTACTTATTTTAAATACAGAAATGCTTTTACCAATGAAATTTTGGCTGTTTATAAT | 3815 | Db | 242476 | TACTTTTGTCCCAAAATAGGAAATCTATTG---TTTTGACTCTTTAGAGATGTTTT | 242421 |
| Db | 243542 | GAGCTTTTAAAGATTTAGATCCCTTTCAGGATTCATGAGTGTCTTCTAGCAAGGGAGATT | 243483 | Qy | 4896 | CCCTTTTAAGTTAAAGAGAGTTTAGTCATTTATTTTGATGCAAAAGTTTAAAGAGAA | 4955 |
| Qy | 3816 | TCTAGAGAACTCAAGTAAACGAGCTAGATTTTATTTATCAAGGAAGTATTTGATAAGTA | 3875 | Db | 242420 | GACTTCTAATATTTTAAATATTAGGAGAAAAATTTCAAGGAGCATATTTTTCATTAGA | 242361 |
| Db | 243482 | AGTTGTGATGGAATACAGGATTAATTAATTGGCCCTGTTAAAGGACATCATCTTTTAAAA | 243423 | Qy | 4956 | GGTTAAATGTTTAAACAAATGTATTCAACAGATTTGTTGGATAATGAGTTCTTGGCCAAT | 5015 |
| Qy | 3876 | TTAGAAAACAATATTTTATTTTGGATATGGAATATCCGAATATTTCAGTTACGGGAACCTGG | 3935 | Db | 242360 | AAATTCATATATGATGTAATGTTCTTTCCTAATATATTTGATCGCTGGTGGTCAGC--T | 242303 |
| Db | 243422 | TCAAATATGCTATATATGACAAAAGTGATTTAACTTCTCTTAATAAGACATGTTGATAG | 243363 | Qy | 5016 | ATTAGAGTCTTATCGAAAAGAAATACGTAGATATCCATTTATTTAAACGCGAAAGATATTT | 5075 |
| Qy | 3936 | CTCGGAAGTCATTCAGGCTATATATCATTTTTTTTATAAATCAGGAATAGTTGGGTTGATT | 3995 | Db | 242302 | ATAAGATATAAATGAACATACAGGCTTTGAAAAATGACTACTTTAGGGGTAAATACCTACA | 242243 |
| Db | 243362 | GTTACAACTAATTTATTTGATAAACAAGGGCTTAAGAAT--AAGAATATTATTCAAAAGA | 243305 | Qy | 5076 | ATCAAGAAAGCAATTTAGTTACGTTGTATTGTGAAATTTTCGCCATAAATATATATGTAAT | 5135 |
| Qy | 3996 | TTACTGATGTTTTCTTTTTTATGTTTATAAAAAAGTTATGAGTTAATGGGAACA | 4055 | Db | 242242 | CTAAG-----AGTGAAGAGGCTGGGTGGAAAAATTTTGTGGGTTTTAGCTATCTT | 242192 |
| Db | 243304 | TTGATGATATAACAATATATCCGAGAAATTTTAACTTCAAAAGAAATTTTATTAACAGGTA | 243245 | Qy | 5136 | GTTATATAAGAAATTTCAAAAGCAGTAGAGGTAAAAATGATAAAAATTTAGTTTATTTGTT | 5195 |
| Qy | 4056 | GCATATTTTATTTTACATCATTAGCCATATTTTTCATATATGAAACAATAGATCCGATT | 4115 | Db | 242191 | TGGAAGGTAGCAAGTTTAAAAAGACGAGTAAAGATGGAATTAATTTCAATAATAGTG | 242132 |
| Db | 243244 | AGGTTGATTTGCTGACTAGTTGTTACCTATTTCTATACATCATTAGAAGGAGTTGGAAA | 243185 | Qy | 5196 | CAAGTTTATAATGTAGATAAATATTAAAGTAGTTGTATAGAAAGCATTTAATCAAAAT | 5255 |
| Qy | 4116 | ATTATATATPAGPACTTTCTTTTCTCAATAGGTATTTGGAATAATATAAATTTTAAA | 4175 | Db | 242131 | CCTGTATATAAATGGAGAAATATATTGGGAGATGCTAGATAGTATTTCTTGAGCAACT | 242072 |
| Db | 243184 | GTTCTCATTTATTTTCAGATTCTCTAAAGATTAGAGTAAGGCTCATAAATTTGATTTTAT | 243125 | Qy | 5256 | TATAAAATATAGAAATATTTTGTAGATAGTGGCTCTGTAGATGATTTCTGTAAATA | 5315 |
| Qy | 4176 | AAGGATATGGAGACAAAATGAATGATTTAAATTTCAAGTTATTTGTACCAATTTTATATGT | 4235 | Db | 242071 | TATCAGAAATTTAGAGATTTATCATATTTGATGATGGCTCTAGTATCCACAGGGGATATT | 242012 |
| Db | 243124 | TTGGATATGGTACTTTATAGAAATGCTTCTAAGGTTTCTTAAAGTTTAAAGAAATAGTTATT | 243065 | Qy | 5316 | TGCAAGGAATATCGAGAAAAGATAAAAGATGTAATAATTTTTTTTCTACTAATCATAGTGA | 5375 |
| Qy | 4236 | CCAAGATTATCTTGATAAATGTTATTAACAGTATTATTAAACCAACATATATAATTTAGA | 4295 | Db | 242011 | TGCGAAAAGTATTTTGTGAAGACAGCGGAATAAATAATTTCTATCAAGAAAACAGAGGC | 241952 |
| Db | 243064 | TGTAATGAAGAGCGAAATGTTATATATTTTACTCTTACATTTTAACAGGGCTACCGA | 243005 | Qy | 5376 | GTATCAAAATGCTAGAAATCATGGAATAAAGCGGAGTACAGCTGAATATATATGTTTGT | 5435 |
| Qy | 4296 | GGTTATTTCTGTAATGATGGAAGTACTGATGATTTGAGAAAATTTGCTTTAAACTATAT | 4355 | Db | 241951 | CAATCAGTTGTCAGCTAACACCGGAGTACTTAGGTGTACAGGAGATTGGAATTTGCTTTT | 241892 |
| Db | 243004 | TTATCCTATC--TGATGATTTCTATGTAATCAGACAAAATAAAATTTTATATGTTGAT | 242946 | Qy | 5436 | GACTCTGATGATGCTGTTGATAGTACTAGTATAGTAAATAATATATTTTAAATATATAAAA | 5495 |
| Qy | 4356 | GAAGAACGATGGAAGAATTAATATTACAGAAAATTAATGCGGCTCAGCAGATGCTCG | 4415 | Db | 241891 | GACATGATGATGTTTTATCTCTTATTCATTTGAGGTTATGTACAAATACAGAAAGCT | 241832 |
| Db | 242945 | TGTTGATGATGGTTCTCAGGACAGTACTAAGGAAATTTGTAAGTAAATTTATATAAAGAA | 242886 | Qy | 5496 | AGTAGAAGTGAATTTATCTGGTTGTTGTAGCTACTTTTTTCAGAAAATATAAATAATTTT | 5555 |
| Qy | 4416 | AAATTTCCGACTAGAACATGCAACAGGTAAATATTTGCTTTTGTGCAATCTGATGACTA | 4475 | Db | 241831 | ACCAATGCCGATATGTTTTAACTAGTATTTGGTAACTTCAACAATACTTACAATACTAGT | 241772 |
| Db | 242885 | TAAAGTTAGTATTTGTTATATCTGTA--TAAACGTTAATGGCGGTAAGCAATTCAGCCTATAA | 242829 | Qy | 5556 | GAAAGTGAATAATCCAAATATTTGATTTTGAAGCAATTAATACCGTGCAGGACATGGGAGAA | 5615 |
| Qy | 4476 | TATAGAAGTTGCAATTTCCGAGAGAAATGCAATGATATAATCACTGAGTATATGCGGATAT | 4535 | Db | 241771 | ATTAATTCACAGTATTTTAAAGAGATTAACCTTTATCTTTAGAGGTGGGTTTGGAGAA | 241712 |
| Db | 242828 | TCTAGCAATGAGATATATGCAACCAAGTGATTTATCATGTATGTTAGATAGTATGATGTTG | 242769 | Qy | 5616 | AAAAATTTTATGAATTTGTATATAAATAATATTTTTTCTACTCTCTGTTTGTAAACTATAT | 5675 |
| Qy | 4536 | AGCAGAGATAGATTTTTTGTGTTTAGTAGCAAAACGGGTATATCAAGAAAAAGAAATAG | 4595 | | | | |

Db 241711 ATGTTATATGGAAAACTTATGGAGTTTCCTGGTTAGCAA-----AATTATATCGAGAA 241657
Qy 5676 AAGAAAAAGATACATAACAGATCTTTTTCAGAGAGAAATCAATGGTTAGGAGAAATTTACTT 5735
Db 241656 GTAAATTTATGAGTAATCCATATCCAGAAAGAAAAATTCATGAAGATATGATCAACACTT 241597
Qy 5736 TTTAATCTGCATTTATTAAGAAATATAGATAGAGTTAGTTATTTTTCAGTGAACATCTTTAT 5795
Db 241596 TTAAGCTTATATCTTGGCTTTCAAAAAATAGCAGTTTGTGTATTTGTCACTGCAGTGGTGT 241537
Qy 5796 TTTTATAGCAGGTATATCAAGTACAGTAATTTCTTTTAAAGAGGTGTCTTTTTCGAA 5855
Db 241536 ATTTTACGTAATATGATCTACTAGGACAAATTCATGAGAGAAATGT-TGTATTTT 241478
Qy 5856 TTGGAATTTGCAAAAACAAGTGTATGTTTAAAGCAAAATATATGTTGAGGATTTT 5915
Db 241477 GAGGCAATTCAAAATAATTTGTATTTTATAAATCTAATATTTTCCACATAATACATCTTTA 241418
Qy 5916 GACGTATCAATTTTAAAGATCTATAGTTGGCAAGTATTTTATATAGCTTACTAATG 5975
Db 241417 ATCTCTGCTGAATATATAATCAAGTATTTGGTGAATAGATATATGTCGGAAGATGATT 241358
Qy 5976 TTTAAATACGGAACAGCTCTATTTTTCACAAATTTTAAATTTTAAAGATCTTTATAA 6035
Db 241357 GATTTTAAATATATGATACAGTTGATTATTTACAGAAAAAATATAGAAAAATTTTAA 241298
Qy 6036 AAATATTTTAACTTGTAAAAAGTATCTAACAAAAATCTTGTCTTAAAAATTTTGT 6095
Db 241297 ACTATCTCTTTAATAATAGATTAGTGTGAAGGAGGTTAATATATATTTATTT 241238
Qy 6096 ATAAGAAATGTTTCGAAACAAAGTTTTHAAAAAATATTTATGGTTATATAGGAAGATATC 6155
Db 241237 TCCTCTATAAGATATTTCACTATTTGTAAGAAAAATA-----TATAATTTGAGGTTATC 241185
Qy 6156 ATGATACTATTTAGTAAATTTCTATATTTGCTACTATATATATATATGTAAGAAAAATTTA 6215
Db 241184 AAGGA-----ATGGAATAAGTTAGTATCATATATCCAGTTTATTAACGTTCAATCTTCTTA 241129
Qy 6216 TCTAAATGTATAGATAGCATTTGTAATCAGACCTTACAAACATATAGAGATTTCTCTGTG 6275
Db 241128 AACGAGTGTATGAAGAGTTTGTAGTCAAACTTATTCAAATTTAGAAATTTATCTAGTT 241069
Qy 6276 AATGACGTAGTACGGATAATTCGGAAGAAATTTGTTTATAGATATCGGAAGAAAGATAGT 6335
Db 241068 AACGATGCTCTACAGATAATTCGGGGATATCTGTGATTATTTTCAGAAATAGATGG 241009
Qy 6336 CGAATTCGTTATTTAAAAAGAGAACGGCGGCTATCAGATGCCGTAATTTATGGCATA 6395
Db 241008 CGTATATTCGTTTTCACAAAAAATAATGTTGTTTATCAGATGCTCGTAATTTATGGAAT 240949
Qy 6396 AGTCGCGCAAGGTGACTACTAGCTTTTATAGACTCAGATGATTTTATTCATTCGGAG 6455
Db 240948 AGTAGACTCTGGGATTATTTATTTAATCTGATTCGGAATGATTTATCTATATAAGAA 240889
Qy 6456 TT---CATCCAAGTTTACAGAACCAATTCAGAGAGAAATGCCCTTTGTGGCAGTTGCT 6512
Db 240888 GATCAATCGAGAGATGTTGTAATTTAGCAGAAATATATTCAGAAATGTTTATAGG 240829
Qy 6513 GGTATG-----ATAGGTATAGCTTCGGGGCAATTTCTTAAACAGACGCGCTTCCT 6566
Db 240828 TGTATGTTGAAAAAAGAGAACACATATTAATAATATCGTTTATAGAGATGAATGATA 240769
Qy 6567 ACAATCAGGCTGTTCTGAGCGGAGGAATGTTTGTAAAGAGCTGCTAGAGCGGATGCT 6626
Db 240768 GAAACTATTTAGTCAGATCAAGCAATACAAATATTTATAA---TTATGATGCTCTATAGA 240712
Qy 6627 CATCGCTTTGTGGCTGTAATAAACTCTATAAAAAAGAACTATTTGAAGATTTTTCGA 6686
Db 240711 GCTATTTTACTGTAGCATAATAAATATATTAAGAGAGAAATTTATTTCTACGCTGTC 240652
Qy 6687 TTTGAAAGGTAAGATTCATGAAGATGAATTTCACTTATCGCTTCTATGAGTTA 6746
Db 240651 TATCCAGTAGGGAAGTTACACGAAGATGAAATTTTAACTTATAAATTTATGAAAGCC 240592

Qy 6747 GAAAAAGTTGCAATAGTAAAGAGTGTCTTGTACTTATTATGTTGACCGAGAAAAATAGTATC 6806
Db 240591 AAAAAATAATTTCTTTTAGATAATAATACTTATGCATATCGTATTAGAGAAAAATAGCATA 240532
Qy 6807 ACAACTTCTAGCAGTACGACCATCGCTTCCTTCCATTCGCTACTGGAATTTCA 6856
Db 240531 ATGACTGGTCTCTACAATATTTAAAAAGGCTACATGCTGTAGAAGCACTTTAA 240482
RESULT 11
AX954530/c
LOCUS AX954530 349980 bp DNA linear PAT 08-JAN-2004
DEFINITION Sequence 1376 from Patent WO03093306.
ACCESSION AX954530
VERSION AX954530.1 GI:40783903
KEYWORDS
SOURCE Streptococcus pyogenes
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1
AUTHORS Telford, J., Massignani, V., margarit y Ros, I., Grandi, G., Fraser, C.
and Tettelin, H.
TITLE Nucleic acids and proteins from streptococcus groups a b
JOURNAL Patent: WO 03093306-A 1376 13-NOV-2003;
Chiron SRL (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
Location/Qualifiers
source
1..349980
/organism="Streptococcus pyogenes"
/mol_type="unassigned DNA"
/db_xref="taxon:1314"
/note="seq 1373 to long, original length:
2.160.266-replaced by-seq 1373: from 0.000.001 to
0.349.980-seq 1374: from 0.300.001 to 0.649.980-seq 1375:
from 0.600.001 to 0.949.980-seq 1376: from 0.900.001 to
1.249.980-seq 1377: from 1.200.001 to 1.549.980-seq 1378:
from 1.500.001 to 1.849.980-seq 1379: from 1.800.001 to
2.149.980-seq 1380: from 2.100.001 to 2.160.266"

ORIGIN

Query Match 11.4%; Score 798.8; DB 6; Length 349980;
Best Local Similarity 48.2%; Pred. No. 9.7e-82;
Matches 3323; Conservative 0; Mismatches 3382; Indels 185; Gaps 31;
Qy 16 TGGCAATTTTGTATATAGTACAGTTCGAATTTCTGCAATCTTAACAAGTCAATATACAA 75
Db 276022 TAGCGATGATTCAAACAGTTGTGGTTTATTTTCTGCAAGTTTGACATTAACATTA 275963
Qy 76 ATGCTGATTTAAATCG--TTCTGGAATTTTATCATAATGATGTTTCATTTATTTGCAAT 132
Db 275962 CTCCCAATTTTAAAGCAATAAGATTTATTTGTTCTTCTATGATACATTTATTTGCT 275903
Qy 133 TTTTATATCTCGPATGCCAGTTGCAATTTGAGTATAGAGTAAATCTGATAGATTTGAAA 192
Db 275902 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 275843
Qy 193 AAACATTTAATATAGTATATATTTTGCATTTTCTTACGGCAGTATCATTTTGTGG 252
Db 275842 TGGTATTTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTTT 275783
Qy 253 AGAATAATTTTCGCACTTTCAAGACGTGTGCGGTATTTTCCACATTTAATAACTTCGTTT 312
Db 275782 AAAACTCTTTTACACACAGCACTTTTCCTTTTACTTTTATGTTCTATGAAATTCGATT 275723
Qy 313 TGGTATACCTATTTTAAACGTAATTTAAGCAGTTTAAAGGATAGCTTTCTATTTTTCGCAA 372
Db 275722 TATTATATCTATTGAATTCATTTTAAATATTATCGAAATATTCTTACGCTAAGTTT 275663
Qy 373 TCTATCAAAAAAGACGATTTCTAATTACACGGCTGAAACGATGGGAAATATGCAAGTTT 432
Db 275662 CACGAGATACCAAGTTGTTTGTGAATACGAATAGGATCTTTTATCAAAAATGACCTTTA 275603

| | | | | | | | |
|----|--------|-----------------------------------------------------------------|--------|----|--------|------------------------------------------------------------------|--------|
| Qy | 433 | TATTTGAATCACATAAACAATTCAAAAAATCTTGTTGCAATCGTAGTTT | 492 | Db | 274531 | TAGAGATTTCTAAGAGAAGAGATTGTATATCATTTGCTCTTTCCCAACAACCGTAATGT | 274472 |
| Db | 275602 | GGAAATAAT---ACGACCAATATATATCGCTGCTGTATCTTGGATTCCTCTGAAAGG | 275546 | Qy | 1547 | CATTAATTTAGTGAATAAATACTATTTCTTAGCTTTCAAAAATTTTACGTGATCAGAAACAG | 1606 |
| Qy | 493 | AAATPAGATAAAATTAATTTATCATACCGCTCTATTATTTCTGTGGAAGAAGCTATAGAGT | 552 | Db | 274471 | CAAAAATTTGGTAAAAAATACTATTTCTAGCTTTTAAAGGTCCTTAGAAAAAGAAAGACAG | 274412 |
| Db | 275545 | ATTGTTATGATTGAAACAATAAATCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG | 275486 | Qy | 1607 | TGTTATTTATTCATCTGGTGGCGCGTGTGCTGCTCCCTTTCTTTTACATCGGAAAACTATT | 1666 |
| Qy | 553 | TTTCAACAGGGAAGTGTGACACAGCTCTTTATAATCTTACCAAGTGAGTTTTAGACG | 612 | Db | 274411 | TGTTATCATATCATCTGGTGGCGCTGTAGCAGTACCATTCTTTATATTTGGAATGTTATT | 274352 |
| Db | 275485 | AGTTAAGCTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCAATTTGAATTTTGGTA | 275426 | Qy | 1667 | TGAGCAAAAGACGATTATATTCGAAGTATTGTATCGAGTTAAATAATCTCATTTAACTGG | 1726 |
| Qy | 613 | TAAAGCAATTC-----GTTTCAGATTTGAGTTGTTAGGTATTGATGTAAGCGTTG | 663 | Db | 274351 | TGTTGTAGAGCGGTTTACATAGAGGTTTTCGACAGATGATTAACCACTTTTGACAG | 274292 |
| Db | 275425 | AATACCAATACAAGATATTATTAATGACATTGAAGCAATGGGAGTGATTTGTCATGTTA | 275366 | Qy | 1727 | AAAACTAGTTTATCCCGTAAACAGATATTTTTTATTTGTTCAGTGGGAAGAAATGAAGAAGT | 1786 |
| Qy | 664 | ATATTAAATTCATTCGGTTTTTACTCGGTTTGAATAAACAATAAATCCAACTGCTAGGTGACC | 723 | Db | 274291 | AAAAATTTAGTGTATCTCTGTAACAGATAAAATTTATTTGTTTCAGTGGGAAGAAATGA | 274232 |
| Db | 275365 | ATGTAGAGGCATTTAGCTTTTGATAATATAGGAGAAAGCGAATCCAACTTTTGAAGGAT | 275306 | Qy | 1787 | ATATCCTTAATCTATTAACTTTGGGGAGTATTTTTTAAATGATTTTTTGTAAACAGTAGGAACT | 1846 |
| Qy | 724 | ATAGCATTTGAACCTTTTCCCAAAATTTTATAAGCCCTAGTCATATCATGATGAACGAC | 783 | Db | 274231 | TTATCCTTAAGGCAATTTAAATTTAGGAGGAATTTTTTAAATGATTTTTTGTACAGTGGGACA | 274172 |
| Db | 275305 | ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT | 275246 | Qy | 1847 | CATGAACAACAGTTTAAATCGATTGATAAAGAGATTGATTATTTGAAAAAATAATGGAAGT | 1906 |
| Qy | 784 | TTTTGGATATACTCGGAGCGTAGTCGGTTTAAATTTTGTGGTATAGTTTCTATTTTGT | 843 | Db | 274171 | CATGAACAGCAGTTCAACCGCTTTATTAAGAAGTTGATAGATTAAAGGACACAGGTCT | 274112 |
| Db | 275245 | TTTTGGATATACCGGTGCTATTATAGGTTTGTCTATGTGGCAATTTGCGCAATTTTTT | 275186 | Qy | 1907 | ATAACCGACGAAATATTTATTTCAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG | 1966 |
| Qy | 844 | TAGTTCCAAATTTATCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAAACGATTGGAC | 903 | Db | 274111 | ATTGATCAAGAAGTGTTCATTTCAACCGGTTTACTCAGACTTTGAACTTCAGAAATTTGTCAG | 274052 |
| Db | 275185 | TAGTTCCCAAAATCAGAAAAGATGGTGGACCGGCTATTTTCTCAAAATAGATAGTCT | 275126 | Qy | 1967 | TATAAAAAATTTCTCAGTTTACAAGAATAAGAAATGGAACAATATATTTAAACAATCAGAAGT | 2026 |
| Qy | 904 | AGAATGAGCGATATTATCATTTACAAAGTTTCGATCGATGATGTTGTTGATGCTGAGGAGC | 963 | Db | 274051 | TGGTCAAAATTTCTCTCATATGATGATATGAACCTTTACATGAAGAAGCTGAGATTGTT | 273992 |
| Db | 275125 | GTAATGGTAGGATTTTAGATTCTATAAATTCAGATCAATCGCAGTAGATGCAACAACA | 275066 | Qy | 2027 | ATTTGCGACGAGCGCCGCTACTTTTATGAAATTCATTATTCAAAAGGAAAAACAATTA | 2086 |
| Qy | 964 | GCAAAAAGACTGCTCAGCCAAACAGATGCAAGGTTGGTATGTTTAAATGGGAA | 1023 | Db | 273991 | ATCACACATGGCGGCCGACGAGCTTTTATGNAATGCGAGTTTCTAAAGGAAAAAACAATTT | 273932 |
| Db | 275065 | TTAAGAAGATTTATTTAGTTCACAATCAATGACAGGG-----CTAATGTTTAAAGTTAGA | 275011 | Qy | 2087 | TTGTTTCTTAGACAAAAAAGTATGTTGAAACATGTAAATGATCATCAAGTAGAGTTTGTGA | 2146 |
| Qy | 1024 | AAACGATCTAGAAATTTACTCCAAATGGACATTTTCATAGCGAAAAACAAGTTTACAGAGT | 1083 | Db | 273931 | GTGGTTCTTAGACACAAGAACAGTTTGGAGACATGTGAATAATCATCAGGTGGACTTTGTT | 273872 |
| Db | 275010 | CGATGATCTAGAAATTTCTAATAATAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT | 274952 | Qy | 2147 | AGAAGAATTTTACAGATAATATATTTATTTATAGAAAATATAGTATGATTTGTTGAA | 2206 |
| Qy | 1084 | TACACAGTTTATTAATGTTTTTAATTTGCGGATATGAGTCTAGTTGGTACAGTCCACTA | 1143 | Db | 273871 | AATAAGGTAAAAACAATGTATAAATTTTGATATCGTTGTAGATATTTGAAGGTTTACAAAAT | 273812 |
| Db | 274951 | TGCTCAATTTCTAATATGTTTTTAAAGGTGATATGAGTTTAGTAGGAACACGCGCTCCCA | 274892 | Qy | 2207 | AAAAATTTGAAGTTCCTAAGCAAACTAACTTTTACATCAAAATTAATAATTT---TTTTTGT | 2263 |
| Qy | 1144 | CAGTTGATGAATTTGAAAAATATATCTCTGCTCAAAAGACAGATTCAGTTTAAACCCAG | 1203 | Db | 273811 | GTAGTCTATGAGGGAATGATGAATCGTCCGTTTTTATAGAACTAATAGTAGTAATTTTATT | 273752 |
| Db | 274891 | CAGTTGATGAATATGAAGAATATTAATTTCAACGCAAGCGACGCTTAGTTTAAAGCCAG | 274832 | Qy | 2264 | GAAAGATTTAAAAACAATAGTTGAAAAATTTTAAATGAGGATCAAGAAAATGAATAATAAAA | 2323 |
| Qy | 1204 | GGATTACAGGTCTCTGGCAGGTTAGTGTGCTAGTAAATATCACAGACTTCGACGACGTAG | 1263 | Db | 273751 | GAAGAAATTTAAGGTAATAATAAAGGAGTGTGCGATGAAAAATCAATAAAAACTCTTTATT | 273692 |
| Db | 274831 | GAATCACTGGTTTGGCAATATCTGGTGAATAATATTAATTAATTAATTTGATGAATCG | 274772 | Qy | 2324 | AGATGCATATTTGATTAATGCTTATCATTAATTTTCTCAGATTTTCTGAGAGGGGATAC | 2383 |
| Qy | 1264 | TTGCGTTGGACTTAGCATATCATGATTAATTTGGAATCTATCTGTCAGATATTTAAATTTTAT | 1323 | Db | 273691 | TTATATTTGCAATATTTTATTTAGTTAAATTTTAAATCACTGGGTTTAGCGGAGGAACTC | 273632 |
| Db | 274771 | TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTGTTGTCAGATATTAAGATTTATC | 274712 | Qy | 2384 | AGATATTTATCATCTCTCTCAGGAGGATGACACCATTTAGTTCTCTCAGAAATACCTGTAT | 2443 |
| Qy | 1324 | TAAAGACGATCAAGATTTGATTTGTTGAGAGGAGCAATGATTAAGTAAAGTAT----- | 1373 | Db | 273631 | AGCTTACAAATATAGT-----ATGTTAGTTGCAAT | 273603 |
| Db | 274711 | TCCTAACACTAAAGGTAGTTTATCTCGGGAACAGGAGCTTAAGTAAAGGTTAGGTTTGAAG | 274652 | Qy | 2444 | AAATATTTTAAATTTCTCAGGATTTATATGTTGAAATTTTACAAAAGATGAGCAAAAATAT | 2503 |
| Qy | 1374 | -----ATGAAAGTTTGTGTTGTTGCTCGTTCTTCAAGGGGACATTTGACTCACTTGTATT | 1426 | Db | 273602 | TTTTACTGTGTGGAATAAAAATTTTATTAGATAGACCTTTATTTTGAAGAAGAAACTCG | 273543 |
| Db | 274651 | GAAATATAATGAATAATTTGTCTGTTGTTTCAAGTGGTGTCTATCAGCACACTTTGAACCT | 274592 | Qy | 2504 | AAAGAAAATAGGATATATGAACGAGTTAAATGTTTACGATTTATTTCTTAATATATACGAA | 2563 |
| Qy | 1427 | GTTAAAAACCGTTTGGGAAGGAAGAACGCTTTTGGGTAAACATTTGATAAAGAGGATGC | 1486 | Db | 273542 | TGATCATCTTTTATTTATTTATTCGCGCACTTTTGAATTTTATTTCT----- | 273498 |
| Db | 274591 | TTTGAACCCATTTGGGAAAAAAGAGATAGGTTTGGGTAACTTTGATTAAGAAAGATGC | 274532 | Qy | 2564 | AAAACTATTGATATGATGTTTGTAGAAATTTTATTAAGAATGTTATCGAGCTTTTGAATAC | 2623 |
| Qy | 1487 | AAGAAGTCTTTTGAAGATGAATAAATGTAATCCATGTTACTTTTCCCAAAATCGCAATCT | 1546 | | | | |

Db 273497 -----TTGTTTCAATAGGTTACTTTTATATTAATTAATCTTTAAATTTTTTTTCTAGCAATTAAGGAC 273442
Qy 2624 TATTACAAAGATTGTTGTTTATTTATGATAGAAATAAAAAACATGGTCTAAGAAATAAGATTG 2683
Db 273441 ATCTCTCAAAAAAGCTTCTCTATAATAATAGGATCGGTAATTTTGGGAGTTCTATTA 273382
Qy 2684 GTTCTAATTGGTTTCGCTTCCACATGATTTTGGGCAATTCCTTTATCAATGAAGACG 2743
Db 273381 AATCAAAATTTTGTCAAAATTAAGTTAATGAATAAATAGTATGTCAATTTTTTATAGGAT 273322
Qy 2744 AAACAGCTTATTTATTTAAAGTAATCTAAATGTCCAGATGAATTTTATATACAGACAATTA 2803
Db 273321 GGACAATTTATCTGAGAGTGA-CTTAGGTTTGGTCACTCACTAATTTTATCAATTT 273263
Qy 2804 TAGAAAAATATGAATTTTCAATAGATTATCTAAATATGAA-----ATTAAAGATATATAA 2860
Db 273262 TTTTGTCTAACTATTTTCTTGTAATATGTACTCAATTTATAACGACTAAAGCCTGTTGT 273203
Qy 2861 AGTGAAAAAATCAACATCTTCTCTATGCTTTTACAGATGATTTCTATTGATGAATTC 2920
Db 273202 GATGGTTTATTTTAAACATTAATTTATTTGATACCAATATCTTTTCAAGGACAGG 273143
Qy 2921 TAAATGCAAGAAATTTAGGTTTATTTTCTAGAAAGTTAAATAGAAATAAATCTA 2980
Db 273142 GTATTATATCGTAATTTTATTTATTTGTAATCTATTTATGTGCAAGAAATAGCTTAATAA 273083
Qy 2981 AA-----TTTAAAGAAATTTACTAAAAATAAAATAGTTGATTTGTGAGAGTAATGTA 3036
Db 273082 AAGAGTATTTATGAATTTAGCACCTATGTACAAATTTTTTTTATTTAGTATTTACCTTTT 273023
Qy 3037 TGTTTAAATTTTAAATATGACCGGGAATTTTATTTTAAAGTACCTTGGTGTGATTA 3096
Db 273022 GAGTTCTACAATTTTTTAAATCAAAATTTGTTCAAAAATTTAGATGTTCTTTTAAACAGG 272963
Qy 3097 TTTTATTCAGACGAAAGTATGTTTATTTTAAATTTTATGAAATTTTATTTTATTTTC 3156
Db 272962 TAGATTACATCTGCTCAATTTACAACTTTGTAGATGGTTTAACTCCTTTTGGAAATAGTTT 272903
Qy 3157 ATATAAAATTTTGAATACTAAGCTAAATTTTAAATAAATGAAATTTTATTTGTTTATTTAT 3216
Db 272902 TAAGAAACAAGTGTCTTATTTGATATAGCTACTCTATGTTTATGATATGTATGTTGT 272843
Qy 3217 GGTCTATATTAATTTGTTTTCAGTGTGCAAGATGATTTTGTGAAATAAATTTTGAAA 3276
Db 272842 AGTACTTACCATGTTTGTATGATAATCTATTTATATCTATAGTAAAAAGA-----TAA 272790
Qy 3277 GATTATTTGAGATTTTACTGCTCCCAATTTTGGATTTATGCAATAATGATTTATAATTT 3336
Db 272789 TCATAAATTGAACCTTCAACTCTCTATTTATAATGCTCTATAATTTATTTACTGAAAGTT 272730
Qy 3337 TGTATTCATTTTATAAATTTGATATAAATAAATAAATAAATAGTATCTTTTATGTTT 3396
Db 272729 TTTATCCAGTGTGGTAATGAATATAGTGGCTAGTTTGTGTAATAATTTTGTGATG 272670
Qy 3397 TAGTTTATTAGGTATATCTGCAATTTGATATATTTTCAAAATGGGAAAGATTTGTAATTTT 3456
Db 272669 GTATCGAACCTTATAAAGAAATTTACTATTTGTGAATAAT--ATATGACATATTTGCTC 272612
Qy 3457 TAGACAGACACTTATAGGACTAGACTATCTTATAACAGGCGTCAAAACAAGTTGGTTG 3516
Db 272611 TGATATGCGAGGAGTAAAGGAAGAAATGATACCTAAAGTTATACATTTATTTGTTGTTT 272552
Qy 3517 GCTTTATGAACTATCTAGTTTAAATACCACTACAAATTTATAGTTTCAATTCGTTAATCT 3576
Db 272551 GGAGAAATCCCTTACCAAGTAATTTTAAAGAAATATATAAATAAATTTGAGAGAAACAATGT 272492
Qy 3577 TTGCACTTTATAAATAAATAAATGCAACAAATTTTTTCTGTGCTCTGCTTTTATACCGA 3636
Db 272491 CCGGATTAAGAAATTTATTTGAATGGAATGAGCAATAATTAATGAT 272448
Qy 3637 TCTATTTAAGTGGATTCGAGAAATTTGGTAGTTTATCGCTAGCAATATTAATATATGCTTGT 3696
Db 272447 --TAGTAAAAATGTTTTATGAGAGAACATATATCTAAGAGAAATTTTGTCTTATGTTTCT 272390

Qy 3697 TATGAGATATATAG--TGGAAAAATTTGCTTTGGATAAAAAAGCTAAATAGTAAATTTTCTA 3755
Db 272389 GACTATGCAAGATTGATATTTATTTATACCTTTATGGGGGTTCTATCTAGATACATGATGT 272330
Qy 3756 ATACTACTTATTTTAAATATCTGAATTTGCTTTACCATGAATTTTGGCTGTTTATAAT 3815
Db 272329 GAGCTTTTAAAAAGTTTATGATCCTTTGAGGATTCATGAGTGTCTTTCTAGCAAGGGAGAT 272270
Qy 3816 TCTAGAGAAATCAAGTAACGAGCTAGATTATTTATTTTCAAGGAAGATTTCATGAAGAATA 3875
Db 272269 AGTTGTGATGTAATACAGGATTAATTAATTTGGCGCTGTTAAAGGACATCACTTTTAAAA 272210
Qy 3876 TTAGAAACAATATTTTATTTTGGATATGGAATATCCGAATATTCAGTTTACCGGAACCTGG 3935
Db 272209 TCAAAATATGCTATATATATGACAAAAGTGAATTAACCTCTCTTAATAAGACATGTGTAGAG 272150
Qy 3936 CTGCGAAGTCAATCAGGCTATATATCTTTTATTAATCAAGGAATAGTTGGGTTGATTT 3995
Db 272149 GTTACAACTAATTTTATTTGATAAACAGAGGCTTAAAGAT--AAGAATATTTATCAAAAGA 272092
Qy 3996 TTACTGATGTTTCTTTTATTTTATATAAAAAAGTTATGAGTTAAATGGGAAAAACA 4055
Db 272091 TTGATGATATAACAATATATCCGAGAAATTTATTTTATCCAGAAATTTTATTAACGGTA 272032
Qy 4056 GCATATTTTATTTTACATCAATAGCCATATTTTTCATATATGAACAATAGATCCGATTT 4115
Db 272031 AGTTGATGTTGCTGACTAGTGTACTTATCTATACATCAATTACGAAGAAAGTTGGAATA 271972
Qy 4116 ATTTATTTATTTAGTACTTATCTTTTCTTCAATAGTATTTGGAATATATAAATTTTAAAA 4175
Db 271971 GTTCTTCAATTTTATTTCTAGATTTCTCTAAAGATTAGAGTAAGGCTCATTAATTCATTTTAT 271912
Qy 4176 AAGGATATGGAGACAAAAAAATGAATGATTTTAAATTTTCACTTATTTGACCAATTTTAAATGT 4235
Db 271911 TTGATATGTTGACTTATAGAAATGCTTTCTAAGGTTTCTAAGTTTAAAGAAATAGTTATTTAT 271852
Qy 4236 CCAAGATTTATCTTGATAAATGATTTTAAACAGTATTTTAAACCAACATATCTAAATTTAGA 4295
Db 271851 TGTAAATAGAAAGGACGAAATGTTATATATTTTACTCTCTACATTTTAAACAGAGGCTACCGA 271792
Qy 4296 GGTATTTCTGTAATATGATGGAAGTACTGATGATTTGAGAAATTTGCTTTAAACTATAT 4355
Db 271791 TTATCTCTATC--TGATGATTTCTCTATGTAATCAGACAAATAAAAATTTTATATGTTGAT 271733
Qy 4356 GAAGAACGATGGAAGAAATTAATATTAAGAAATTTAATCGCGCTCTAGCAGATGCTCG 4415
Db 271732 TGTTGATGATGTTTCTGAGGACAGTACTAAGGAAATTTGTAAGTAAATTTATATAAAGAGAA 271673
Qy 4416 AAATTTCCGACTAGAACATGCAACAGGTAATAATATTTGCTTTTGTGCTGATTTCTGATGCTA 4475
Db 271672 TAAAGTTAGTATTTGATATCTGTA--TAAACGTAATGGCGGTAAAGCAATCAGCCTATAA 271616
Qy 4476 TATAGAGTTGCAATGTTTCGAGAGAAATGCAATGATATAATAAATGAGTATAATGCGGATAT 4535
Db 271615 TCTAGCAATGAGATATATGCAACCAAGTGAATATCATGTATGTGTAGATAGTGAATG 271556
Qy 4536 AGCAGAGATAGATTTTGTTTTGTAGAGAAACCGGTATACAAAGAAAAAAGAAATAG 4595
Db 271555 GTTATTAGAGATGCAAGTTGAATTTATTTTAAAGATTTTAGAGAGCCTTAACGTTGCTCAA 271496
Qy 4596 TAAATTTTCACTCTTAACGAGAGAAAGACGTGTAAGAAATTTTGTGTCAGGATCTTAATAT 4655
Db 271495 TAGATATGTAGGCTT-----TGTTTATCTCGATATTTCAATTAACCAAGGTAATTAATG 271442
Qy 4656 AGAAAAATAATTTTGGTGCAGCTTTATTTACAGAGATATTTATAAAGATATAAATTTCCA 4715
Db 271441 GTTGAATCCCAAAATATTAGAAGTTTAAATTTTCTGATTTAAAAATATAAATATCACTTAAA 271382
Qy 4716 AATTAAATATAGAGTATTTGTTGAGGATTTGCTTTTAAATTTTGGAGGCTTTGAACAATGT 4775
Db 271381 AATTGAAACTTGTATTGTTTAAATAA--TGCTTATTTAGTAGATTTTCGAAATTTCTCTTGT 271324

| | | | |
|---------------------------------------------------------------------------------------------------------|--------|--------------------------------------------------------------------|--------|
| Qy | 4776 | AACAGCTGATGTAGTGGATCATTAGAGAAATTAATTTAATAATGATGCATTCGTAAACAGTTC | 4835 |
| Db | 271323 | TTTCAAGGGGAGAAATTTCTTATCAGAAGAAAATAATGATATATATTTATCAAAAAAGGGA | 271264 |
| Qy | 4836 | GCTTTATTAATCAGAAATCTCTATATAATTAATTAATGATTTAGTCCAAAGTTGGAGAATTA | 4895 |
| Db | 271263 | TACTTTTGTCCCCAAATAGGAAATCTATTG-----TTTGACTACTTTAGGAAGATGGTTT | 271208 |
| Qy | 4896 | CCCCTTTAAAGTTAAAAAGAGAGTTTAGTCAATTTATTTTGATGCCAAAGTTATTAAGAGAA | 4955 |
| Db | 271207 | GACTTCTAATATTTTTTAATTTATGGGAAAAAATTTCAAAGGACTATTTTTTCATTTAGA | 271148 |
| Qy | 4956 | GGTTAAATGTTTAAACAAAATGTATTCACAAGATTTGTTGGATAATAGTTCCTTGCCAAT | 5015 |
| Db | 271147 | AAATTCATATATGATGTAAATGCTTTTCTCTAATATATTTTGATCGCTGGTGCAGC--T | 271090 |
| Qy | 5016 | ATTAGAGTCTTATCGGAAAGAAATACGTAGATATCCATTTATTAAGCGAAAGATATTT | 5075 |
| Db | 271089 | ATAAGATAAAAATGAACATACAGGCTTTTGAANAATGACTCTTTTAGGGGTAAATACCTACA | 271030 |
| Qy | 5076 | ATCAAGAAAGCATTTTAGTTACGTGTGATTTTGATGAAATTTTTCGCTAAACTATATGTAAT | 5135 |
| Db | 271029 | CTAAG-----AGTGAAGAGCTGGGTGGAANAATTTTGGTGGTTTAGCTATCTT | 270979 |
| Qy | 5136 | GTTATATAAGAAATTTCAAAGCAGTACGAGGTAAAAAATGGATAAATTAGTGTATTGTT | 5195 |
| Db | 270978 | TGCAAGGTAGCAAGGTTTAAAAAGAGCGAGTAAAGAAATTAATTTCAATAATAGTG | 270919 |
| Qy | 5196 | CGAGTTTAAATGTAGATAAATTTTAAGTAGTTGTTAGTATAGAAAGCAATTTTAANTCAAAT | 5255 |
| Db | 270918 | CCTGTATATAATGGAGAAATATATATTTGGGAGATGCTAGATAGTATTTCTTGAGCAAACT | 270859 |
| Qy | 5256 | TATAAAAATATAGAAATATTTTGTATAGATGATGCTCTGTAGATGATTTCTGCTAAATA | 5315 |
| Db | 270858 | TATCAGAAATTGAGATTAATCAATTTGATGATGCTCTAGTGATCCGACAGGGGATAT | 270799 |
| Qy | 5316 | TGCAAGGAATATGCAGAAAAAGATAAAGAGTAAAAAATTTTTTCACTAAATCATAGTGA | 5375 |
| Db | 270798 | TGCGAAAGTATTTTTTGGAGACAGCGGAAATAAATATTTCTATCAAGAAAAAGAGGC | 270739 |
| Qy | 5376 | GTATCAAAATGCTAGAAATCATGGAATTAAGCGGAGTACAGCTGGAATATATTTATGTTT | 5435 |
| Db | 270738 | CAATCAGTTGCGAGTAAACAACGGAGTACTTAGTGTACAGGAGATGGATTTGCTTTTTA | 270679 |
| Qy | 5436 | GACTCTGATGATGTTGTTGATAGTAGATTTAGTAGAAAAATATATTTTAATTTATAAAA | 5495 |
| Db | 270678 | GACAGTGATGATGTTTATCTCTCTTATTCAAATGAGGTATGTACAAATACAGAAAGCT | 270619 |
| Qy | 5496 | AGTAGAAGTGATTTATCTGGTGTGTTTGACGCTACTTTTTTCAGAAAAATATAAATAATTTT | 5555 |
| Db | 270618 | ACCAATGCCGATATGTTTAACTAGTATTTGTTAATCTTCAACATCTTACATACTAGT | 270559 |
| Qy | 5556 | GAAGTGAATAATCCAAATATGATTTTGAAGCAATTAATACCGTGCAGGACATCGGAGAA | 5615 |
| Db | 270558 | ATTAAATTCAGTATTTTAAAGAGATTTAACTTTTATCTTTTAGAGGTGGCGTTCGAAGAA | 270499 |
| Qy | 5616 | ABAAATTTTATGAATTTGTATATAATAATTTTTTCTACTCTCTGTTTGAACATATAT | 5675 |
| Db | 270498 | ATGTATTTAGGAAAACTTATGGAGTTTCCCGTTAGCAA-----AATATATCCGAGAA | 270444 |
| Qy | 5676 | AAGAAAAGATACATAACAGATCTTTTTCAGAGAAATCAATGTTTAGGAGAAAGATTTACTT | 5735 |
| Db | 270443 | GTAAATTTTGTAGTATCCATATCCAGAGGAAAAATTCATGAAGATATGATACAACTT | 270384 |
| Qy | 5736 | TTTTAATCTGCATTTTAAAGAAATAPAGATAGAGTTAGTTATTTTGACTGAACATCTTAT | 5795 |
| Db | 270383 | TTAAGCTTATCTTCGCTTCAAAAATAGCAGTTTGTGATATTTGCTACTGCAGTGGTGT | 270324 |
| Qy | 5796 | TTTTTATAGGAGGATATCTAAGTACAGTAAATTTCTTTTAAAGAGGTGCTTTTTTGCAA | 5855 |
| Db | 270323 | ATTTTAGCGATATAATGATCTACTAGGACAAAATTTCAATGAGAGAAATGT-TGTATTTTTT | 270265 |
| Qy | 5856 | TTGAAAAATTTGCAAAAAACAGTGTATGATTTTGTTTTAAAGCAAAATATATGTTGAGATTTT | 5915 |
| Db | 270264 | GAGGCAATTTCAAAAATAATTTGTTATTTATAAATCTAAATTTTCCACATAAACATCTTTTA | 270208 |
| Qy | 5916 | GACGTATCAATTTGTTTAAAGATACTACGTTTGGCAAGTATTTTATTATAGCTTACTAATG | 5975 |
| Db | 270204 | ATCTCTGCTGTTAATATATAATGAAGTATTTGGTGAATAGATATATGTTGGGAAGATGAT | 270145 |
| Qy | 5976 | TTTAAATACGGAACACAGTCTATTTTTCACAAAATTTTAAATTTTGAATACTTTTATAAAA | 6035 |
| Db | 270144 | GATTTTAAATATATATGATACAGTTGATTTATTCAGAAAAAATATAGAAAAATTTTAAAG | 270085 |
| Qy | 6036 | AAATATATTTTAACTGTTTAAAGTATCTAACAAAAATCTTTTGTCTTAAAAATTTTTTGT | 6095 |
| Db | 270084 | ACTATTTCTTTTAAATAGGATTTAGTGTGAAGGAGGTTAAATATATATATATTTATTTAT | 270025 |
| Qy | 6096 | ATAAGNAATGTTTCCGACAAAGTTTAAAAAATAATTTATGTTTATATAGRAAGATATC | 6155 |
| Db | 270024 | TCCTCTATAAGATATTTTCACTATTTGTAAGAAAAATA-----TATATTTTGGAGTTATC | 269972 |
| Qy | 6156 | ATGGATACTTATAGTAAAAATTTCTATAATTTGACTATATATATAATGTAGAAAAATTTTA | 6215 |
| Db | 269971 | AAAG-----ATGGATAAGTTAGTATCATTTACCAGTTTATACGTTTCAATCTTCTTA | 269916 |
| Qy | 6216 | TCTAAATGTATAGATPAGCAITTTGTAATCAGACCTCACAAACATATAGAGATTTCTCTGGT | 6275 |
| Db | 269915 | AACGAGTGTATAGAAGTGTTTTAGCTCAAACTTATTCAAAATTTAGAAAATTTATCTAGTT | 269856 |
| Qy | 6276 | AATGACGTAGTACGGATAATTCGGAAGAAATTTGTTTAGCATATGCGAAGAAAGATAGT | 6335 |
| Db | 269855 | AACGATGGCTACAGATAATTTCCGGGGATATCTGTGATTTATTCAGAAAATAGATGGG | 269796 |
| Qy | 6336 | CCCATTTGCTTATTTTAAAAAAGAGAACGGCGGGCTATCAGATGCCCGTAAATTTATGGCATA | 6395 |
| Db | 269795 | CGTATTTGCTTTTTTCAAAAAATATGCTGTTTATCAGATGCTCGTAATTTATGGAATTT | 269736 |
| Qy | 6396 | AGTCGCGCAAGGGTGACTACTTGTAGCTTTTATAGACTCAGATGATTTTATTCATTCGGAG | 6455 |
| Db | 269735 | AGTAGAGCTACTGGCGATTTATTTTATTTACTTGTATCGGATGATTTATCTATAAAGAA | 269676 |
| Qy | 6456 | TT---CATCCAAAGTTTACACGAGCAATTCAGAGAGAGATGCCCTTGGCGAGTTGCT | 6512 |
| Db | 269675 | GATGCAATCGAGAGAAATGGTTGAATTTTAGCGAGAAATAATAATTCAGAAAATTTTATAGG | 269616 |
| Qy | 6513 | GGTTATG-----ATAGGTAGATGCTTTCGGGGCATTTCTTAAACAGACGAGCGCTTCTCT | 6566 |
| Db | 269615 | TGTTATGTTGAAAAAGAGAACACATATTTAAATATCGTTTGAAGATGGAATGATA | 269556 |
| Qy | 6567 | ACAAATCAGGCTGTTCTGAGCGCGAGGAATGTTTGTAAAAAGCTGTAGAGCGCGATGGT | 6626 |
| Db | 269555 | GAAACTTATTTAGTCCAGTACAAAGCAATACAAAATATTTTATAA--TTATGATGCCTATAGA | 269499 |
| Qy | 6627 | CATCGCTTTGCTGGCTGTAAATAAACTCTATAAAAAGAACTATTTTGAAGATTTTCGA | 6686 |
| Db | 269498 | GCTATTTTTTACTGTAGCACATAATAAATTTATTAAGAGAGAAATTTTCTTCTACGCTGTC | 269439 |
| Qy | 6687 | TTTTGAAAAGGGTAAAGATTCATGAAGATCAATACTTTCCTTATCGTCTCTCTATGAGTTA | 6746 |
| Db | 269438 | TATCCAGTAGGGAGGTATACAGAGATGAATTTTAACTATAAATATATTTTGAAGGCC | 269379 |
| Qy | 6747 | GAAAAAGTTGCAATATAGTTTAAAGAGTGTCTGTACTATTTATTTGTGACCGAGAAAAATAGTATC | 6806 |
| Db | 269378 | AAAAATATAATTTTCTTTTAGATAATAACTTATGCATATCGTATTAGAGAAATATAGCATA | 269319 |
| Qy | 6807 | ACAACTTCTAGCATGACGTGACCATCGCTTCCATTCGCTTACTGGAATTTCA | 6856 |
| Db | 269318 | ATGACTGGTTCTCAATAATTTAAAGGGTACATGCTGTAGAAGCACTTAA | 269269 |
| RESULT 12 | | | |
| AF355776 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| Streptococcus agalactiae CNCT 1/82 type IV capsular polysaccharide synthesis operon, complete sequence. | | | |
| linear BCT 29-APR-2001 | | | |

| | | | | |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-------------|--|
| ACCESSION | AF355776 | AF355776.1 | GI:13876772 | |
| VERSION | | | | |
| KEYWORDS | | | | |
| SOURCE | Streptococcus agalactiae | | | |
| ORGANISM | Streptococcus agalactiae | | | |
| | Bacteria; Firmicutes; Lactobacilliales; Streptococcaceae; Streptococcus. | | | |
| REFERENCE | 1. (bases 1 to 17596) | | | |
| AUTHORS | McKinnon, K., Chaffin, D.O. and Rubens, C.E. | | | |
| TITLE | Direct Submission | | | |
| JOURNAL | Submitted (03-MAR-2001) Department of Infectious Disease, Immunology and Rheumatology, Children's Hospital and Regional Medical Center, 4800 Sand Point Way NE, Seattle, WA 98105, USA | | | |
| FEATURES | Location/Qualifiers | | | |
| source | 1. 17596 | | | |
| | /organism="Streptococcus agalactiae" | | | |
| | /mol_type="genomic DNA" | | | |
| | /strain="CDC 1/82" | | | |
| | /db_xref="taxon:1311" | | | |
| gene | 817..17187 | | | |
| -35_signal | /gene="type IV capsular polysaccharide synthesis operon" | | | |
| -10_signal | 817..822 | | | |
| | /gene="type IV capsular polysaccharide synthesis operon" | | | |
| gene | 834..839 | | | |
| -35_signal | /gene="type IV capsular polysaccharide synthesis operon" | | | |
| -10_signal | 928..2396 | | | |
| gene | /gene="cpsA" | | | |
| RBS | 928..933 | | | |
| CDS | /gene="cpsA" | | | |
| | 939..2396 | | | |
| | /gene="cpsA" | | | |
| | /note="similar to Ltyr" | | | |
| | /codon_start=1 | | | |
| | /transl_table=1 | | | |
| | /product="CpsA" | | | |
| | /protein_id="AAK43602.1" | | | |
| | /db_xref="GI:13876770" | | | |
| | /translation="MSNRRSRQKKHSHTPLRVINFLVIFILLVSVLSFLMYRHHF LAFNLNVIYGVWTVLIIILASFLICIKNKAFTTIIILVASIFVATVLYGFKSTIDL TNNLKATSYSEIEMSVVVPKDSKITNLEAVSKLAPVKNDSNTIDLIHKKSEKGI SITQKTSYQDAYNRIKNGDSQAMLNNAVSLIELSTPDPFKQIKTIYTKKKKI ARKNTKHGQYDIYISGIDTFGSIIVTSVSDVNIIMVTNTHKVLTTTPRDAYVK IPDGGGQYDKLHAGLYGVTSMTKLENDLYDINLYVARNFSPFLKIDLLGGVTV YNDQATFKHGNFPPVPGQVTLNSEQALGFVRERYSLQGGNDRGRNOKVIAAIINK LASSQVTKLNSITLSQTSVQTNTIDINILNQLSTQRTFVESQALTGHSTG ELPSYMPGAQLYMSIDQSSLSNAKSKIKNTMEE" | | | |
| gene | 2388..3133 | | | |
| RBS | /gene="cpsB" | | | |
| CDS | 2388..2393 | | | |
| | /gene="cpsB" | | | |
| | 2402..3133 | | | |
| | /gene="cpsB" | | | |
| | /codon_start=1 | | | |
| | /transl_table=1 | | | |
| | /product="tyrosine phosphatase CpsB" | | | |
| | /protein_id="AAK43603.1" | | | |
| | /db_xref="GI:13876771" | | | |
| | /translation="MIDHSHIVFDVDDGPKTLBSLSLIESYRQGVRIIVSTSHRR KMPETPDDIIFKNPSIVHAEAKRFEHLQILYGGELYITSDMLEKLLKQIPILNNT KPALTEFSMTQSWDKIDHPSNLVLMGLITPVVAHRYNALENKERVKEIINMGCVT QINSHILKQKLFNDKRRFKGRARYFLEENLVHFVASDMENLDVRPPFLAEAYKIIC RDFGERANQLFIENAQSILKNHYI" | | | |
| gene | 3132..3834 | | | |
| RBS | /gene="cpsC" | | | |
| CDS | 3132..3137 | | | |
| | /gene="cpsC" | | | |
| | 3142..3834 | | | |
| | /gene="cpsC" | | | |
| | /codon_start=1 | | | |
| | /transl_table=1 | | | |
| | /product="putative capsular polysaccharide chain length regulator Cpsc" | | | |
| | /protein_id="AAK43604.1" | | | |
| | /db_xref="GI:13876772" | | | |
| | /translation="MNKIANTVEINIFNLKKLWKKKFLITFVAIAFATAGLFYSLF IVPQYTSSTRIYVINPNTNNSITAOLOAGSFLANDYKEIITSTVDLEKVISSEKL NYPSSQLQKITISILKDTVISISVEDANPKMSQKLANVREAAVSKIKAVTOVEDI TTLEGNLPKAPSPNKKVNLIGFIVGAGLSTIVLIMGILDRVNTTEEDIEKVLGL TSIGIVPDLNKL" | | | |
| gene | 3833..4543 | | | |
| RBS | /gene="cpsD" | | | |
| CDS | 3833..3839 | | | |
| | /gene="cpsD" | | | |
| | 3845..4543 | | | |
| | /gene="cpsD" | | | |
| | /codon_start=1 | | | |
| | /transl_table=1 | | | |
| | /product="capsular polysaccharide chain length regulator/exporter CpsD" | | | |
| | /protein_id="AAK43605.1" | | | |
| | /db_xref="GI:13876773" | | | |
| | /translation="WTRLEIVDSKLRQAKTTEEFNAIRTNIQSGKENKILATSVR EGEKSTTSISLALSQAQFKTLIDADTRNSVMSGTFKATGTYLGNLNYLSGNADL GDITCEYNVRLMVPVPSKVPNPNTALLQNAVFNKMEIAIKNIIDFYIITDTPPIGLVV DAATIANACDGFILVTQAGRIKRNVEKAKEQSGSKFLGIILNKVNESVATYGDY GDYGNCKRDRKRK" | | | |
| gene | 4535..5944 | | | |
| RBS | /gene="cpsE" | | | |
| CDS | 4535..4540 | | | |
| | /gene="cpsE" | | | |
| | 4595..5944 | | | |
| | /gene="cpsE" | | | |
| | /codon_start=1 | | | |
| | /transl_table=1 | | | |
| | /product="glucose-1-phosphate transferase CpsE" | | | |
| | /protein_id="AAK43606.1" | | | |
| | /db_xref="GI:13876774" | | | |
| | /translation="MIOTVVVFSASLTTLITPNFKSNKDLLFVLHIVFVLSDF YRDFWRSGLIEEFKFWLKYSPYIFISSLFFFKNSFTTTRLSFFFTITNLSILLYL LNSFLKYRYKSYAKFRSDTKVILITNKDSLKRMFTKNDHNTIACILDSSERDYL DLKHSNRIKNDALTSCLTVDQAFINIPIELFGKYQIDIINIEAMGVIVNVN VLEASPNIGEKRIQTPEGYSVTYSMKFYSHLIAKRFLDITGAILGLLICGVAI FLVPIQIRKGGPAIFSONRVGRNGRIIPRFYKFSRMRVDAEOIKDILLVHNTQMTGLMPK LEDDPRTKTKGPIRKTSIDELPQFYNVLKDSLSLVGTTRPTVEYKYNSTQKRRLS FKPGITGLWQISGRNNITDFDEIVKLDVQYINWSIWSDIKIILLTLKVLVLGTGAK" | | | |
| gene | 5958..6417 | | | |
| RBS | /gene="cpsF" | | | |
| CDS | 5958..5963 | | | |
| | /gene="cpsF" | | | |
| | 5968..6417 | | | |
| | /gene="cpsF" | | | |
| | /codon_start=1 | | | |
| | /transl_table=1 | | | |
| | /product="beta-1,4-galactosyltransferase enhancer CpsF" | | | |
| | /protein_id="AAK43607.1" | | | |
| | /db_xref="GI:13876775" | | | |
| | /translation="MKICLVSSGSHLAHLNLLKPILEKEDRFWTFDKEDARSILRE EIVVHCFPDRNNRVNLTILAPKVLKRPDVIISGAAVAVPPFYIGKLFCKT VYIEVDFRDKTPTLTGKLVVPTDKFIVQWEMKKVYPKAINLGIP" | | | |
| gene | 6402..6908 | | | |
| RBS | /gene="cpsIVG" | | | |
| CDS | 6402..6407 | | | |
| | /gene="cpsIVG" | | | |
| | 6417..6908 | | | |
| | /gene="cpsIVG" | | | |
| | /codon_start=1 | | | |
| | /transl_table=1 | | | |
| | /product="beta-1,4-galactosyltransferase CpsIVG" | | | |
| | /protein_id="AAK43608.1" | | | |
| | /db_xref="GI:13876776" | | | |
| | /translation="WIFVTVTGHEQFNRLIKEVRLKGTDAIDQEVFIOTGYSDPEP QNCQWSKFLSYDDMNSYKAEAEIVITHGGPATFMNAVSKGKTIIVPRQEQFGHVN HQVDFVANKVTIMNFIDIVDIERLQNVVYEGTMMRPLETNRSNFIREFKVLKELCD ENQ" | | | |
| gene | 6882..7992 | | | |
| | /gene="cpsIVH" | | | |

| | | | | | |
|------|-------------------------------------------------|----|------|------------------------------------------------------------------|------|
| RBS | 6882. .6887 | Db | 4830 | AAACCTCTTTTAAACGACGACGACTTTCCTTTTACTTTTATCTATGAAATTCGATTT | 4889 |
| CDS | /gene="cpsIVH" | Qy | 313 | TGGTATACCTATTAAAGTAATTAATTAAGCAGTTTAAGGATAGCTTTCTATTTTCGACAA | 372 |
| | 6895. .7992 | Db | 4890 | TATTATATCTATTGAATTCATTTTAAATAATTATCGAAAAATTTCTTACGCTAAGTTTT | 4949 |
| | /transl_table=11 | Qy | 373 | TCATCAAAAAAGAGCAGTTCTAATTACACGGCTGAACGATGGGAAAAATATGCAAGTTT | 432 |
| | /product="putative polyaccharide repeating unit | Db | 4950 | CACGAGATACCAAGTTGTTTTCATACGAATAAGGATTCCTTTATCAAAAAATGACCTTTA | 5009 |
| | /polymerase CpsIVH" | Qy | 433 | TATTTGAAACACATAACAATAATCAAAAAATCTTGTTCATTTGGTAGTTTGGTACAG | 492 |
| | /protein_id="AAK43609.1" | Db | 5010 | GGAAATAA---ACGACCATATAATATATCGCTCTGTATCTTGGATTCCTCTGAAAGG | 5066 |
| | /db_xref="GI:13876777" | Qy | 493 | AAATAGATAAAATAAATTTATCAATACCGCTCTATTATTCTGTGGAGAAAGCTATAGAGT | 552 |
| RBS | 7875. .7880 | Db | 5067 | ATTGTTATGATTTGAAACATAACTCGTTAAGGATAATAACAAGATGCTCTTACTTCAG | 5126 |
| gene | /gene="cpsIVH" | Qy | 553 | TTTCAACAAGGGAAGTGGCGACCGCTTTTATAAATCTACCAAGTAGTTTTTAGACG | 612 |
| CDS | 7949. .8800 | Db | 5127 | AGTTAACCTGCTTAACTGTTGATCAAGCTTTTATAACATACCAATTTGAATTTATTTGGTA | 5186 |
| | /gene="cpsIVH" | Qy | 613 | TAAAGCAATTC-----GTTTCAGATTTTTCAGTTTGTAGTTTGTAGTTATGATGAGCGTTG | 663 |
| | /codon_start=1 | Db | 5187 | AATACCAAAATACAAGATAATTATTAATGACATTCGAAGCAATGGGAGTGAATGTCAATGTTA | 5246 |
| | /transl_table=11 | Qy | 664 | ATATTAAATTCATTCGGTTCGCTGAAAAAACAACCAAAATCCCACTCGTAGGTGACC | 723 |
| | /product="putative glycosyltransferase CpsIVH" | Db | 5247 | ATCTAGGACCTAGCTTTGATAATATAGGAGAAAGCGAATCCAAACTTTTGAAGGAT | 5306 |
| | /protein_id="AAK43610.1" | Qy | 724 | ATAGCATTTGTAATTTTCCACAAATTTTATAAGCCTAGTCATATCATGATGAAACGAC | 783 |
| | /db_xref="GI:13876778" | Db | 5307 | ATAGTGTATTACATATCTTCAAAATTTCTATAAATATAGTCACCTTATATGCAAAAACGAT | 5366 |
| gene | 8931. .9680 | Qy | 784 | TTTTGGATATACCTCGGACGGTAGTCGGGTAAATTAATTTTGGCTATAGTTTCTATTTTGT | 843 |
| RBS | /gene="cpsIVH" | Db | 5367 | TTTTGGATATACCGGGTGTCTATTATAGGTTTGTCTCATATGTCGATTTGTGGCAATTTTC | 5426 |
| CDS | 8931. .8936 | Qy | 844 | TAGTCCCAATTAATTCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC | 903 |
| | 8943. .9680 | Db | 5427 | TAGTCCCAAAATCAGAAAAGATGGTGGACCGGCTATCTTTTCTCAAAATAGAGTAGTCTC | 5486 |
| | /codon_start=1 | Qy | 904 | AGAATGGACGCATATTACATCTCAAGTTTCGATCGATGTATGTTGATGCTGAGGAGC | 963 |
| | /transl_table=11 | Db | 5487 | GTAATGGTAGGATTTTGTAGATTTCTATAAATTCAGATCAATCGAGTAGATGCGAACAACAA | 5546 |
| | /product="putative glycosyltransferase CpsIVH" | Qy | 964 | GCAAAAAGACTTGTCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTAAATGGGAA | 1023 |
| | /protein_id="AAK43611.1" | Db | 5547 | TTAAGAAAGATTTTATTAGTTTCACAATCAATGACAGGG-----CTAATGTTTAAAGTTAGA | 5601 |
| gene | /db_xref="GI:13876779" | Qy | 1024 | AAACGATCCTAGAAATTAATCTCAATTCGACATTTTATACACGCAAAAACAAGTTTAAACGAGT | 1083 |
| | /transl_table=11 | Db | 5602 | AGATGATCCTAGAAATTAATTAATAATAGGAAATTTATTTCG-AAAAACAAGCATAGATGAGT | 5660 |
| | /product="putative glycosyltransferase CpsIVH" | Qy | 1084 | TACCAAGTTTATAATGTTTAAATTTGGCGATATGAGTCTAGTTGGTGTACAGCTCCACCTA | 1143 |
| | /protein_id="AAK43612.1" | Db | 5661 | TGCCTCAATTTCTATAATGTTTAAAGGTGATATGAGTTTGTAGGAAACAACGCGCTCCCA | 5720 |
| | /db_xref="GI:13876780" | Qy | 1144 | CAGTTGATGAAATTTGAAAAATATACCTCTGCTGCTCAAAAGAGACGATTCAGTTTAAACAG | 1203 |
| | /transl_table=11 | Db | 5721 | CAGTTGATGAAATTAAGAAAGTATAATTTCAACGAGAAAGCGACGCTTAGTTTAAAGCAG | 5780 |
| | /product="putative glycosyltransferase CpsIVH" | Qy | 1204 | GAATTAACAGTCTCTGGCAGGTTAGTGGTCTGATGTAATATACAGACTTCGACGACCTAG | 1263 |
| | /protein_id="AAK43613.1" | Db | 5781 | GAATCACTGGTTTGTGCAAAATATCTGGTAGAAATATAATTAATGATTTTGTAGTAATCG | 5840 |
| | /db_xref="GI:13876781" | Qy | 1264 | TTCCGTTGCACTTAGCATTAATTAATTTGGAATCTCTGGTCAATATTAATTAATTAATTTAT | 1323 |
| | /transl_table=11 | Db | 5841 | TAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTTGGTCAATATTAAGATTTATTC | 5900 |
| | /product="putative glycosyltransferase CpsIVH" | Qy | 1324 | TAAAGACAGTGAAGTTGTTATTTGTCGAGAGAGGAAAGTAAAGCTA----- | 1372 |

Db 5901 TCCTAAACACTAAGGTAGTCTTACTCTGGGACAGGTGCTAAGTAAGGTAAAGTTTGAAG 5960
Qy 1373 -----TATGAAGTTTGTGGTGGTCTTCCAGGGGACATTTGACTCCTCTGTATTT 1426
Db 5961 GAATATAATGAATTTGCTCTGGTTGGTTCAAGTGGTGTCTAGCACACTTTGAACCT 6020
Qy 1427 GTTAAACCGTTTGGGAAGGAAGAAGACGTTTGGTGAACATTTGATGAAGAGGATGC 1486
Db 6021 TTTGAACCCCATTTTGGAAAAGAAGATAGGTTTGGGTAAACCTTTGATGAAGAAGATGC 6080
Qy 1487 AAGAAGTCTTTTGAAGAAATGAATAAATGTATCCATGTTTCTCAACCAATCGCAATCT 1546
Db 6081 TAGAGTATCTAAGAGAAGATGTTATATCATGTTCTTTCAACCAACCGTAATGT 6140
Qy 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTCAAAATTTTACGTGATGAGAACCAGA 1606
Db 6141 CAAAACCTTGTAAAAAATACTATTCTAGCTTTTAAAGTCTTAGAAGAAGAACCCAGA 6200
Qy 1607 TGTATTAATTCATCTGGTGGCGGTTGCTGTGCCCTTCTTTTACATCGGAAACATTT 1666
Db 6201 TGTATCATATCATCTGGTGGCGGTTAGCAGTACCATTTTATATTTGTTAAGTTATTT 6260
Qy 1667 TGGAGCAAGACGATTATATTTGAAGTATTTGATCGAGTTAATAAATCTACATTAACCTG 1726
Db 6261 TGGTTGTAAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACCACTTTGACAG 6320
Qy 1727 AAAACCTAGTTATCCCGTAACAGATATTTTATTTGTTCCAGTGGGAAGAAATGAAGAAGGT 1786
Db 6321 AAAATTAGTGTATCTGTAAACAGATAATTTATTTGTTCCAGTGGGAAGAAATGAAGAAGGT 6380
Qy 1787 ATATCCTAAATCTATTAACTTTGGGAGTATTTTATTAATGATTTTGTAAACAGTAGGAAT 1846
Db 6381 TTATCCTAAGCAATTAATTTAGGAGGAATTTTATTAATGATTTTGTCCACAGTGGGACA 6440
Qy 1847 CATGAACAAAGTTAATCGATTGATAAAGAGATGATTTTATGAATAAATAAATGGAAGT 1906
Db 6441 CATGAACAGCAGTTCAACCGTCTTATTAAGAAGTTGATAGATTAAAGGGACAGATGCT 6500
Qy 1907 ATAAACGACGAAATATTTTATTAACACAGATATTTCTGACTATTTCCAGATAATTTGCAAG 1966
Db 6501 ATTGATCAAGAAGTGTCTCATTAACCGGTTACTCAGACTTTGAACCTCAGAAATGTGCG 6560
Qy 1967 TATAAAAAATTTCTCAGTTACAAGAAATGGAACAATATATTAACAATCAGAAAGTAGTT 2026
Db 6561 TGGTCAAAAATTTCTCATATGATGATGAATGAATCTTATCATGAAGAAGCTGAGATGTT 6620
Qy 2027 ATTTGCCAGGAGGCCCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAACAATTA 2086
Db 6621 ATCAACATGCGGTCACGACGCTTTATGAATGCAGTTTCTTAAAGGGAAAAAATCTATT 6680
Qy 2087 TTGTTTCTAGACAAAAAAGTATGTCGAACATGTAATGATCATCAAGTAGAGTTTGTA 2146
Db 6681 GTGGTCTCTAGACAAGAACAGTTTGGAGAGCATGTAATATCATCAGGTGGAATTTGTT 6740
Qy 2147 AGAAGAAATTTTACAAGATAATAATTTTATTAAGAAATATAGATGATTTGTTGAA 2206
Db 6741 AATAAGGTAAACAAATGATAATTTTGATATCGTTGTAGATATTTGAAGGTACAAAAT 6800
Qy 2207 AAAATTAATGAGTTTCTAAGCAAACTAACTTTTACATCAAAATAA---TAAATTTTGT 2263
Db 6801 GTAGTCTATGAGGGAACGATGAATCGTCCGTTTTTGAAGAACTAAACAGAAATTTTATT 6860
Qy 2264 GAAGATTTAAACAAATAGTTGAAAAATTTAATGAGGATCAAGAAATGAATAATAAAA 2323
Db 6861 GAAGAAATTTAAGGTAATTAAGAGGTTGTGTAGTGAATAATCAATAAAAACTCTTTATT 6920
Qy 2324 AGATGCATATTTGATAATGGCTTATCAATTTTCTCAGATTTTCTAGATTTTCTGAGAGGATAC 2383
Db 6921 TTATATGCAATTTTATGTTAAATTTTAAATCACTAGGTTTAGGAGGGGAGCTC 6980
Qy 2384 AGATATTAATCATCTCTCAGGAGAAATGCACCACTTAGTTTCTTCCATCAGAAATACCTGTAT 2443
Db 6981 AACTTACAAATAGTGATGTTTGTGCAATCTTCTTGTGTGGAATAAATTTTATTAGA 7040

Qy 2444 AATTAATTTTAAATTAATCTCAGGATTTATATATGTTGAATTTTACAAAAGATGAGCAAAAATAT 2503
Db 7041 TAGCCTTTATTTGAAAGAGAAAACCTCGTTATCATCTCTTTTATTTATTTATTTGCGCCAT 7100
Qy 2504 AAAGAAAATAGGATATATGAACGAGTTTAAATGTTTACAGATTTATTTCTTAATATATCAGAA 2563
Db 7101 TTTGAATTTATCTTTCTGTCATAAGGTTACTTTTATATTAATTAACCTTTAAATTTTCTTAGC 7160
Qy 2564 AAAACTATTGATATGATGTTTGAAGATTTTATTAAGAATGATCGAGCTTTTGAATPAC 2623
Db 7161 ATTA--AAGGATATCTCTCAAAAAAGCTTCTCTATAATAATAGGATCCGATTTTGT 7218
Qy 2624 TATTTACAAAGATGTTGTTTATTTGATAGATAAAAAACATGGCTCTAAGAATAGATTTG 2683
Db 7219 GGAGTTCTATTAATCAAAATTTTGTGAAATTAGATTTAATAGAAAATTAATAATATCAAT 7278
Qy 2684 GTTCTAATTTGGTTCCTCCACATGATTTTGTGGCAATTTCTTTTATCAAAATGAAGAAG 2743
Db 7279 TTTTATAGGAGTGGACAAATTTATTTCTGAGAAGTGAATTTAGGTTTGTGTCATCCCTAATCT 7338
Qy 2744 AAACAGCTTATTTTAAAGTAATCTAAATGTCAGATGAACATTTTATATACAGACAAATTA 2803
Db 7339 ATTCTAATTTTGTGAGTAACCTGTTTATATATGTAACATTTTATATAGAAACTA 7398
Qy 2804 TAGAAAAATATGAATTTTCAAAATAGATTTATCTAAATATGGAATTTAAGATATATAAAGT 2863
Db 7399 AGATTAATACTATTGCTTTTATTTAACTCTAAATTTACTTCTGTATCAGTACTATTAT 7458
Qy 2864 GGAATAAATCAACATCTTCTTATGTTCTTACAGATGATCTAT--TGATGAATTGCTA 2922
Db 7459 TCAAGAACTGATATATATAGTACTCTTATTTATATCTATTTATATATGTTTACAAAGAT 7518
Qy 2923 AATCAAGAAATTTAGGTTTATTTGCTAGAAAGTTTAAAAATAGAAAAATAATCTAAA 2982
Db 7519 AACCTGATAGGAATAATTTTATGATAGTTGCTCCGTACATCAACTGTTCTTGTAGCA 7578
Qy 2983 TTTTAAAGAAATTTACTAAAAATAAATAAGTTTGTGAGAGTAATGATGTTTA 3042
Db 7579 TTTACTTTCTTCTGCTCTACTATTTTCTCAACTCAAAATTTGTTCAAAAATTAGATAGC 7638
Qy 3043 AATTAATTAATATGACCCGGAATATTTATTTTAAAGTACTTCTGTTGATTTATTTA 3102
Db 7639 CTTTGTGACAGTAGTTTAAACTGCTCATTTACAGCTTTGAGACGGCTTAACTCTTTT 7698
Qy 3103 TTCCAGAGCAAAAGTATGTTTATTTTATTAATTTTATGAAATTAATTTTCTCATATA 3162
Db 7699 GGAATAGTTTAAAGGAGACGAGTGTCTTATTTGATTAATAGTACTCTATGTTATTGAGT 7758
Qy 3163 AATTTTGAACCTAAAGCTAATAATAAAAAATGAAATTTTATTTTATTTATGCTCTA 3222
Db 7759 ATGTTATGTTAGTACTTTCCATGTTTGTATGATAATCTATATCTATAGTAAAAA 7818
Qy 3223 TATTAATGTTTGTTCAGTAGTCAAGATAGTTTGTGTAATGAAATAAATTTTGAAGATTTAT 3282
Db 7819 GTCAATGTAGTTGAGTCCAGATAC-----TTTTGTTTATAATGCTATAGTATTTATTTA 7873
Qy 3283 TTGAGATTTTACTGCTCCCATATTTGGATATTTGCAATTAATGTTATTTATTTGTTAT 3342
Db 7874 CAGAGAGTTTATCCCAAGTATAGTTATGAATTTAGTTGGATGGTTTTCGGAAAAAT 7933
Qy 3343 CATTTATAAATATTGATTATAAAAAATTTAAAAATAGTATCTTTTATAGT--TTTTTATG 3400
Db 7934 TTTGTGGGGGTAGATGATTTTACACGAGAGTTCTCTTGACGGCAATAAATAATAGT 7993
Qy 3401 TTTTATAGGTATCTGCAATTTGATATTTATTTCAAAATGGGAAAGATATTTGTTTATAGA 3460
Db 7994 GTAATGTACAGTATATAATTTGCAACAAATTTTATAGTCTGCGTTGATTTCAATTAGA 8053
Qy 3461 CAGACCTTATAGGACTAGACTATCTTTAACAAGGCGTCAAAACAAGGTTGGTTGGCTT 3520
Db 8054 AAACAAACATATAAGAAATTTGGAATTTATCTTGTTTATGATGGATCAACAGATGGTAT 8113

| | | | | | | | |
|----|------|------------------------------------------------------------------|------|----|-------|--------------------------------------------------------------------|-------|
| QY | 3521 | TATGAACCTACCTACGTTAAATACCACCTACAAATATATAGTTTCAATCCGTTATCTTTGC | 3580 | Db | 9189 | GGGATATATTTAGATACAGATGTTGAACCTTATATAAACCTATTGACGACCTTTTAAATGTT | 9248 |
| Db | 8114 | AAAGAGTTATGTGAGGAGATAGAAAATCAGATGAAAGAAATTTAGACATTTTCAAAAACA | 8173 | QY | 4657 | GAAAAATAATGTTTTGGTCAAGCTTTTATTTCAAGAGATATATAAAAGATATAAAATCCAA | 4716 |
| QY | 3581 | ACTTATAAAAATAAAATGCAACAAATTTTTTTTCTTGTTGTTGTTTATATACCGATCT | 3639 | Db | 9249 | GAAAGTTATTTTCTGTTGAGCT--ACCTGGAGAGGTTAATACAGGTTTGGGTTTGGGA | 9305 |
| Db | 8174 | AATGGAGGACAAATCAAGCGCAAGGAAATTTAGGTATTTTATATCTACAGGAGATTTGATT | 8233 | QY | 4717 | ATTAAATATAGAGTATTTGGTGAGGATTTCTCTTTTAAATTTGGAGGTTCTTGAAACAATGTA | 4776 |
| QY | 3640 | ATTAAAGTGTATCGAGAAATTTGGTATTTATCGCTAGCAATATTA--TTATATGCTTGT | 3696 | Db | 9306 | TCAATTTTCAGAGAAATCTTTTATAAAGAAATAATGAAAAATCTACGAAG-ATACAGAATA | 9364 |
| Db | 8234 | GGTTTTTTGACAGCGACGATACAAATTTGACCTTAAAAATGATGAAACGTTTACTAAATATA | 8293 | QY | 4777 | ACAGGTGTAGTGTGATACTAGAGAAATATTTATATAATTTATGTCATTCGTTAAACAGTTCTG | 4836 |
| QY | 3697 | TATGGAGATATATAGGTGGGAAAAATTTCTCTGGATATAAAAAAGCTAATAGTAATATTTGTAA | 3756 | Db | 9365 | CTATAGTTTAAAAAAAACGTGTGGCAATTTACCCAGATTTTACTAACAGAGCAGGCT | 9424 |
| Db | 8294 | TATGAAGATGAACAGTAGACTGGGTGCAATGTAATCACAATAAAATTTTACTCTAACCGGT | 8353 | QY | 4837 | CTTTATTAAATCAGAAATTTCTCTATAAATAATTTGATTTAGTCAAGATTTGGAGAAATTAC | 4896 |
| QY | 3757 | TACTACTTATTTTAAATTAATCGAATTTGCTTTTACCATGAAATTTTGGCTGTTTATAAAT | 3816 | Db | 9425 | TAAGTACATAAAATAAAATTTCAAGTAAATAAAGCGTATTTTTTAATTTATCTCTCAAGTTATTT | 9484 |
| Db | 8354 | GTTAACTTTATATATAATGACCTGAATACTATAATGTCTTAAATAAACAAGATTTTCTTA | 8413 | QY | 4897 | CCCTTTAAGTTAAAAAGAGAGTTTAGTCATTTATTTTGTATGCCAAAAAGTTTATTAAGAGAA | 4956 |
| QY | 3817 | CTAGAGAAATCAAGTAACGAAGCTAGATTTATTTATTAAGGAAGTATTTGATAAAGTAT | 3876 | Db | 9485 | TTGTCAGTAAATATGTTTAACAAAATAAAATAAATAAATAAAGAGAGAAACATTTAACGATTC | 9544 |
| Db | 8414 | TACGAATTTCTGAGTACAAATAAGATTTTGTAGTTCACTCTCGAGGGGTTTGTATCTAGA | 8473 | QY | 4957 | GTTAAATGTTTAAACAATAATGTTTCAACAGATTTGTTTGGATTAATGATTTCTTCCCAATA | 5016 |
| QY | 3877 | TAGAAAAACAATATTTTATTTGGATATGGAATATCCGAATATTCAGTTTACGGGAACTTGGC | 3936 | Db | 9545 | CAATTTAGGAGGAAATTTGGGAGGAGAAAAAATAAATTTGATTTAGGTAGATAGATAAAGTT | 9604 |
| Db | 8474 | GATTTAGCTTTAAAAATAAAATTCCTGGAGAAAAAATAAATAAATGAAATACACAGTTTAT | 8533 | QY | 5017 | TTAGAGTCTTTATCGAAAGAAATACGTAGATATCCATTTTATTTAAAGCGAAAGATATTTTA | 5076 |
| QY | 3937 | TCGGAAGTCAATCAGGCTATATATCATTTTATTTTAAATCAGGAATAGTTGGGTGATTT | 3996 | Db | 9605 | AGGGTTTCAITTTACTACTAGATAAAGATATTTGGATATGGAACGTACAGAAATATTTTAA | 9664 |
| Db | 8534 | TTTGATCTCAATAAAATGCTAATAAGTTTGTATTTATTAAGCCACCTTTTATTAATTAC | 8593 | QY | 5077 | TCAGAAAGCATTTAGTTAGCTTGTGTAATTTGATGAAATTTTTCGCCTAAACTATATATGTAATG | 5136 |
| QY | 3997 | TACTGATGTTTCTTTTATTTGTTATAAATAAAAGTTATGAGTTAAATGGGGAACAG | 4056 | Db | 9665 | GGCAGTAAAAAATAAGTAAAGGTATCAAGGTTAACTGATTTTCAITTTTAAATAGTAACTCT | 9724 |
| Db | 8594 | TACTACAGAAAAAATAGTACAAACAACCTTCTCATATAGTAGTATCAATGGGCATAATC | 8653 | QY | 5137 | TTATATAAGAAATTTCCAAAAGCAGTAGAGTAAATAATGGAATAATTTAGTGTATTTCTTC | 5196 |
| QY | 4057 | CACATTTTATTTTACATCAATAGCCATATTTTTCATATATGAACAATAGATCCGATTA | 4116 | Db | 9725 | TGTCAACTAATAATTTATAGGAAGTAAATGGAATAATTTTAAATAGTATATATTTTC | 9784 |
| Db | 8654 | GATATCTGTACTGAGTGTATTTATATGAAAGGATTTTAA--ATGGATTTGGAAGAGTTG | 8711 | QY | 5197 | CAGTTTATAATGTAGATAAATATTTAAGTAGTTTGTATAGAAAGCATTTATTTAACTCAAAAT | 5256 |
| QY | 4117 | TTATATATTTAGTACTATTTCTTTTCAATAGGTATTTGGAATAATATATAATTTTAAAA | 4176 | Db | 9785 | CGATATATAATGTAGAAAAATTTCTTAGAGGAATGTAATGATAGTTTATTAATCAAACTT | 9844 |
| Db | 8712 | CTTTTCAAGATTTATTTGGTGCAATTTTCGTTAGTAGCTAATAAATTTGTATATAATAAG | 8771 | QY | 5257 | ATAAAAATATAGAAAAATTTATTCATATAGATGATGGCTCTGTAGATGATCTGCTGCTAAAAATAT | 5316 |
| QY | 4177 | AGGATATGGAGACAAAAAATGAATGATTTAAATTTCAGTATTTGTACCAATTTATAATGTC | 4236 | Db | 9845 | ATAAAAATATAGAAAAATCTGTTAGTAGATGATGGTTCAACGGATAAATTCAGGAATTTATCT | 9904 |
| Db | 8772 | ATTATAGAAAAACCGAAGAAATTAAGATAATAGCTTATTTTTTCAAGAAAAAATTTATG-- | 8829 | QY | 5317 | GCAAGGAATATGCAGAAAAAGATAAAGAGTAAATAATTTTTTTCACCTAACTCATAGTCGAG | 5376 |
| QY | 4237 | CAAGATTTATCTGATAAATGTTTAAACAGTATTTAACCACCAACATATCTAATTTTAGAG | 4296 | Db | 9905 | GTGATTAATTTAGCCTGAAGAAGATAAAGCAATCAGGGTATTTTCAATAGNATANTGGAGGTT | 9964 |
| Db | 8830 | AAAATTTTAAAAATAGATTTTATTTGGTCATAGTCGAAAACTTTCTGTGATAGTTTCTTAA | 8889 | QY | 5377 | TATCAAAATCTAGAAAAATCATGGAATAAAGCGGAGTACAGCTGAATATATTTATGTTGTTG | 5436 |
| QY | 4297 | GTTATTTCTCGTAAATGATGGAAGTACTGATGATTTCTGAGAAAAATTTGCTTTAAACTATATG | 4356 | Db | 9965 | TATCCGATGCTCGGAATTTTGGTGTAGTAAATCTCGAAGCGCGGTATATTTCTTTTATTTG | 10024 |
| Db | 8890 | TTTTTTTCCATTTCTATACAAAAATTTTATTAGAAAAATATAGAGGAGAGTACAAATGATAG | 8949 | QY | 5437 | ACTCTGATGATGTTGTTGATAGTATAGTAAAAAATTTATATTTTAAATATTTATAAAAA | 5496 |
| QY | 4357 | AAGAACGATCGAAGAAATTAATATTAACAAGAAATTTAATGCGCGGTCTAGCAGATGCTCGA | 4416 | Db | 10025 | ATTTCAGACGATTTATTTGATAAAAAATTTATTTAGAAAAATGTTATTTGTTTAAAGTAATC | 10084 |
| Db | 8950 | AAAAAACTATGATACCTTAAAAAATAACATTTATTTGTTGGTGGTGAAT--CCTAAGAGT | 9008 | QY | 5497 | GTAGAAGTGAATTTATCTGTTGTTTCTGACGTACTTTTTCAGAAAAATATAAATAAATTTTG | 5556 |
| QY | 4417 | AAITTCGGACTAGAACATGCAACAGGTAATAATATTTGCTTTTCTCGATTTCTGATGACTAT | 4476 | Db | 10085 | ATAAGGTAGATATGGTTATCTGCAATTTATTTGAGTGTTT-----ATAAATAATG | 10132 |
| Db | 9009 | GAACTGTTATTTAAATGTATCAATCATGGGAAAAAATTTTGTCCAGATTTACGAATAATTT | 9068 | QY | 5557 | AAGTGAATAATCCAAATATTTGATTTTGAAGCAATTAATACCGTCAGGACATGGGAGAAA | 5616 |
| QY | 4477 | ATAGAAGTTGCAATGTTTCGAGAGAAATGCATGATAATAATACTGAGTATATATGCGGATATA | 4536 | Db | 10133 | ACTTGAACCGATCGTTAGTCGTTTACAGATTTGTATATATATTTCTAGGGTGGAAATTAT | 10192 |
| Db | 9069 | GAATGGAATGAACAAAAATTTATGATGTCATATAAATTTACATATACTAGACAAAGCTTATAA | 9128 | QY | 5617 | AAAAATTTTATGAATTTGTATATAAATAATTTTTTTCTACTCTCTGTTTGTAAACTATATA | 5676 |
| QY | 4537 | GCAGAGATAGATTTTGTGTTAGTAGAGAAAAACGGGTATACAAAGAAAAAAGAAATAGT | 4596 | Db | 10193 | TTAGTAAATTTATATGTTAAATATAAGGATCTCTTTTACACAGCCTCGGGGAAGTTTATAA | 10252 |
| Db | 9129 | GAAGGAAAGTATGCTTTTCGTGCTGATTTATGCTAGATTTGATATTTATTTATCAATATGTT | 9188 | QY | 5677 | AGAAAAAGATACATAACAGATCTTTTTTCAAGAGAAATCAATGGTTAGGAGAGAGATTTACTTT | 5736 |

| | | | | | |
|-------------|-------|----------------------------------------------------------------|-------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Db | 10253 | GAACAGATATCGCTAAGAAGTAACATTTCCAAATTTGGTAAACTACACAGGATGAATTTA | 10312 | CDS | /gene="cpsiaE" 352..1527 /gene="cpsiaG" /codon_start=1 /transl_table=11 /product="CpsiaE" /protein_id="AAL57060.1" /db_xref="GI:18033321" |
| Qy | 5737 | TTAATCTGCATTATTAAAGAAATAGATAGAGTTAGTTATTGTACTGAACATCTTTATT | 5796 | | |
| Db | 10313 | CAACTTATAATATTACTTATATTCAGATAAATAGCCTTTGTACCAGAACCACTATATT | 10372 | | |
| Qy | 5797 | TTTATAGAGAGGTATATCAAGTACAGTAATCTTT-TTAAAGAAGGTGTGTTTGGCAA | 5855 | | |
| Db | 10373 | TTTATAGAAGGAGAGAGGCTAGTATAATAGCTTCTTCTTACAGTAAAGAAATTTAGATT | 10432 | | |
| Qy | 5856 | TTGGAATTTTGCRAAACAAGTGTATGTTTGAAGCAATATATGTCAGGATTTT | 5915 | | |
| Db | 10433 | CTTTAGAAGCTTATGAGAAGCGATTAAATATTGGAAGAAATATATATCTATATCAG | 10492 | | |
| Qy | 5916 | GACGTATCAATTTGTTAAAGATACTATACGTTGGCAAGTATTTTATATAGCTTACTAATG | 5975 | gene | /gene="cpsiaF" 1551..2000 /gene="cpsiaF" /codon_start=1 /transl_table=11 /product="CpsiaF" /protein_id="AAL57061.1" /db_xref="GI:18033322" |
| Db | 10493 | AACTGTGTACATATACCTTAGTTTAACTTATATCATATATACAGATTTCGTAATATC | 10552 | CDS | |
| Qy | 5976 | TTTAAATACGGAACAGCTCTATTTTGCAGAAATTTTAAATTTTTPAGAAATCTTTATAAA | 6035 | | |
| Db | 10553 | CTGAGGTATGCAATAAACAAGAAATATTGGATAAGTTTAACTATTTTAAAGTATA | 10612 | | |
| Qy | 6036 | AAATATATTTTAACTCTTTAAAGTATCTAACAAAATTTCTTCTCTTAAATTTTGT | 6095 | | |
| Db | 10613 | AAAATATTTTAGGATTAAATAATAGACTAAATAATTTATACATTAAAGAAATTTAGGCATAC | 10672 | | |
| Qy | 6096 | ATAAGATTTGTTTCAACAAAGTTTTTTAAAGAAATATTATGCTTATATAGCAAGATAT | 6154 | gene | /gene="cpsiaG" 2000..2482 /gene="cpsiaG" /codon_start=1 /transl_table=11 /product="CpsiaG" /protein_id="AAL57062.1" /db_xref="GI:18033323" |
| Db | 10673 | CTTATCAAGATATAACATAAATCTTGAACTGCAAGAAATTAAGAAATGAAGAAAT | 10731 | CDS | |
| RESULT 13 | | | | | |
| AF332894 | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | 4513 bp DNA linear BCT 07-JAN-2002 |
| ACCESSION | | | | | Streptococcus agalactiae strain NCDC S815 capsular polysaccharide |
| VERSION | | | | | bioynthesis gene cluster, partial sequence. |
| KEYWORDS | | | | | AF332894.1 GI:18033319 |
| SOURCE | | | | | Streptococcus agalactiae |
| ORGANISM | | | | | Streptococcus agalactiae |
| REFERENCE | | | | | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; |
| AUTHORS | | | | | Streptococcus. |
| TITLE | | | | | 1 (bases 1 to 4513) |
| JOURNAL | | | | | Kong, F., Gowan, S., Martin, D., James, G. and Gilbert, G.L. |
| MEDLINE | | | | | Streptococcus. |
| PUBMED | | | | | 21635022 |
| AUTHORS | | | | | 11773119 |
| TITLE | | | | | 2 (bases 1 to 4513) |
| JOURNAL | | | | | Kong, F. and Gilbert, G.L. |
| FEATURES | | | | | Direct Submission |
| Source | | | | | Submitted (02-JAN-2001) CIDMLS, ICPMR, Westmead Hospital, Darcy |
| gene | | | | | Rd., Sydney, NSW 2145, Australia |
| CDS | | | | | Location/Qualifiers |
| | | | | | 1..4513 |
| | | | | | /organism="Streptococcus agalactiae" |
| | | | | | /mol_type="genomic DNA" |
| | | | | | /strain="NCDC S815" |
| | | | | | /serotype="Ia" |
| | | | | | /db_xref="taxon:1311" |
| | | | | | <1..126 |
| | | | | | /gene="cpsiaD" |
| | | | | | <1..126 |
| | | | | | /gene="cpsiaD" |
| | | | | | /codon_start=1 |
| | | | | | /transl_table=11 |
| | | | | | /product="CpsiaD" |
| | | | | | /protein_id="AAL57059.1" |
| | | | | | /db_xref="GI:18033320" |
| | | | | | /translation="AKEQMEQSGSKFLGIILNKVSVSATYGDYGNKRDKRK" |
| 352..1527 | | | | | |
| gene | | | | | |
| Query Match | | | | | 10.4%; Score 729.4; DB 1; Length 4513; |
| ORIGIN | | | | | |
| | | | | | LNLSFKYRKYSAKFSRDTKVLI1TNKDSLSKMTFRNKYDHNVIACVILDSSEKDCY |
| | | | | | DLKHSRI1NKDALTSCLTVDQAFIN1PIELFGKYQI0D1INDIEANGV1VNVN |
| | | | | | VEALSFNIEKRIQTFEGSVIYYSMKFYKYSHLIAKRFLDITGAILLIGLIGLVAI |
| | | | | | FLVQIRKDGPAIFSQNRVGRNRI1FRFYKSRMRVDAEQIKDOLLVHNMQTMGLMF |
| | | | | | LDDPRITKIKGIRKTSIDELPQFVNLKGDMSLVGRPTPTVDEYKYNQTKERLS |
| | | | | | FKPGITGLWIGISGRNITDFDEIVKLDVQYINWSIWSIDIKI1ILLTLKVVLLTGAK" |
| | | | | | 1551..2000 |
| | | | | | /gene="cpsiaF" |
| | | | | | 1551..2000 |
| | | | | | /gene="cpsiaF" |
| | | | | | /codon_start=1 |
| | | | | | /transl_table=11 |
| | | | | | /product="CpsiaF" |
| | | | | | /protein_id="AAL57061.1" |
| | | | | | /db_xref="GI:18033322" |
| | | | | | /translation="MKICLVSSGGHIAHLNLLKPIWEKEDRFWVTFDKEDARSILRE |
| | | | | | EIVVCFPPNKNVKNLVKNTILAFKVLKRPDVI1ISSGAAVA1VPFYIGKLGCKT |
| | | | | | VYIVRDIRIDKPTLTGKLVVPTDKFVQWEMKKVYPKAINLGGIF" |
| | | | | | 2000..2482 |
| | | | | | /gene="cpsiaG" |
| | | | | | 2000..2482 |
| | | | | | /gene="cpsiaG" |
| | | | | | /codon_start=1 |
| | | | | | /transl_table=11 |
| | | | | | /product="CpsiaG" |
| | | | | | /protein_id="AAL57062.1" |
| | | | | | /db_xref="GI:18033323" |
| | | | | | /translation="MIFVTGTHQFNRLIKEVDRLKGTGAIDQEVFIQTYSDPEP |
| | | | | | QNCQSKFLSYDDMNSYKEAEI1VTHGGPATFNNVSKGKKT1VPRQEQFGHVN |
| | | | | | HOVDKFLKFLKELDYILNISELENI1KEKNVISTKVISQNDPFCFSFKNELSKLE |
| | | | | | " |
| | | | | | 2578..3717 |
| | | | | | /gene="cpsiaH" |
| | | | | | 2578..3717 |
| | | | | | /gene="cpsiaH" |
| | | | | | /codon_start=1 |
| | | | | | /transl_table=11 |
| | | | | | /product="CpsiaH" |
| | | | | | /protein_id="AAL57063.1" |
| | | | | | /db_xref="GI:18033324" |
| | | | | | /translation="MQILLILALIVLLICSSYNEKMKFLNMAEIFFVFIYIILTSIL |
| | | | | | LHSLFKNPDPRIILAAENSLIGVSLAFVHYKNTLELKLKAPLNGLLIFELG |
| | | | | | GTYYLCHNNIIONISIFGRDLIGSDWNGMHTQAMGFFESNLI1PIYVTVNL1YIYI |
| | | | | | YMKLRNYSIMTVGLVLLFTFILPGSGRAGIVAILAQMFILLNLTVVVVKRTIKFL |
| | | | | | YILPFLVIMVLYFDNLLSIYIRI1NLRSGSESRSFLYKDYNTVINNLLFQEGV |
| | | | | | KEIWNLDPLGSHSTYIGYFSGLLGMNIVPGLLI1FTNIGKAKQSAFYFIEVG |
| | | | | | TLTILFSFPALEDLDGANWLIVFTFVLGILENKFYSQLKRWKS" |
| | | | | | 3717..4513 |
| | | | | | /gene="cpsiaI" |
| | | | | | 3717..4513 |
| | | | | | /gene="cpsiaI" |
| | | | | | /codon_start=1 |
| | | | | | /transl_table=11 |
| | | | | | /product="CpsiaI" |
| | | | | | /protein_id="AAL57064.1" |
| | | | | | /db_xref="GI:18033325" |
| | | | | | /translation="MEKRILVSI1PIYNSAYLKECVSQSVLQOQTHPLIEVILIDGGS |
| | | | | | TDSGEICDNLSDQRNRLVFKHNGGVSSARNLGDSTGEFTTFVDSDDFVAPNMI |
| | | | | | EIMLNLI1TENADIAEVDEDI1SNERYDKKKRNFYKFKNNNSIKFLSGNRVENIV |
| | | | | | CTKLYKKSII1GNLRFDENLKGEDLLFNCKLLCOEHRIVVDVTTSSLYTYRIVKTSAMN |
| | | | | | QKFNESLDFITTFNFSVSSLVPAKLANYVEAKFUREKIKLURKMFELG" |

| | | | |
|---------------------------------------------------------------------------------------------------------------------|------|-------------------------------------------------------------------|------|
| Best Local Similarity 59.2%; Pred. No. 2.6e-73; Matches 1390; Conservative 0; Mismatches 916; Indels 41; Gaps 7; | | | |
| Qy | 16 | TGGCAATTAATTGATATAGTAGCAGTTGGCAATTTCTGCAATCTTAAACAAGTCATATACCAA | 75 |
| Db | 173 | TAGCGATGATTCAAACAGTTGTGGTTATTTTCTGCAAGTTTGACATTAACATAATTA | 232 |
| Qy | 76 | ATGCTGATTTAAATCG--TTCTCGAATTTTATCATAAATGATGGTTTCATTTATTTTCAT | 132 |
| Db | 233 | CTCCCAATTTTAAAGCAATAAGATTTATTTGTTGTTCTATGATACATTTATTTGTCT | 292 |
| Qy | 133 | TTTTTATATCTCGTATGCGAGTTGCAATTTGAGTATAGAGGTAATCTGATAGAGTTTCAAA | 192 |
| Db | 293 | TTTATCTTTCTGATTTTACAGAGACTTTTGGAGTCTGGCTATCTTTGAAGAGTTTAAAA | 352 |
| Qy | 193 | AAACATTTAACTATAGTATATAATTTTGCATTTTTCACGCGAGTATCAATTTTGTGTGG | 252 |
| Db | 353 | TGGTATTTCAATAACAGCTTTTACTATATTTTCAATCAAGTTTCATTTATTTTATTTTAA | 412 |
| Qy | 253 | AGATATAATTCGCACTTTCAAGACGTGTGCGGTATTTTTCACATTAATAAATCTCGTTT | 312 |
| Db | 413 | AAAACTCATTTACACGACGACTTTCCCTTTTCTTTTATTTGCTATGAATTCGAATTT | 472 |
| Qy | 313 | TGGTATACCTATTAAACGTAATTAATTAAGCAGTTTAAAGGATAGCTTCTATTTTCGACAA | 372 |
| Db | 473 | TATTTGATCTATTCGAATTCATTTTAAATATATATCGAAATATTTCTTACGCTAAGTTT | 532 |
| Qy | 373 | TCATATAAAAGAACGATCTCAATTAACAAGCTGAAACGATCGAAGGAAATATGCAAGTTT | 432 |
| Db | 533 | CACGAGATACCAAGTTGTTTGTATACGAATAAGGATCTTTTATCAAAAATGACCTTTA | 592 |
| Qy | 433 | TATTTGATACATAAACAATTCAAAATAATCTTTGTCATTTGATGTTTGTAGGTACAG | 492 |
| Db | 593 | GGAATAAAT---ACGACCATTAATTAATTTGCTGTCTGCAATCTTGACCTCTCGAANAAG | 649 |
| Qy | 493 | AAATAGATAAAATTAATTTATCATTTACCGCTCTATTATTTCTGTGGAAGAGCTATAGAT | 552 |
| Db | 650 | ATTGTTATGATTTGAAACATAACTCGTTAAGGATAATAACAAGATGCTCTTACTTCAG | 709 |
| Qy | 553 | TTTCAACAAGGAAGTGGTCGACACGCTGTTTATAAATCTACCAAGTGAGTTTGTAGACG | 612 |
| Db | 710 | AGTTAACTGCTTAACCTGATCAAGCGTTTATTAACATACCCATTGAAATTAATTTGGTA | 769 |
| Qy | 613 | TAAAGCAATTC-----GTTTCAGATTTTGAATTTGTTAGTTATGATGATAGCGTTG | 663 |
| Db | 770 | AATACCAAAATACAAGATATTAATTAATGACATTTGAAGCAATGGGAGTGATTTGCAATGTA | 829 |
| Qy | 664 | ATATTAATTCATTCGGTTTACTGCTTGAATAAATAAATAAATAAATAAATAAATAAATAA | 723 |
| Db | 830 | ATGTTAGGCACTTAGCTTTGATTAATATAGAGAAAGCGAATCCAACTTTTGAAGGAT | 889 |
| Qy | 724 | ATAGCATTTGAATCTTTTCCAAATTTTATAAGCCCTAGTCATATCATGATGAACGAC | 783 |
| Db | 890 | ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTACACCTTATAGCAAAAACGAT | 949 |
| Qy | 784 | TTTTTGGATATCTCGGAGCGGTAGTCTGGTTAAATTTTGGGTATAGTTTCTATTTTGT | 843 |
| Db | 950 | TTTTTGGATATCACGGGTGCTATTATAGGTTTGTCTATATGTCATATGTCGCAATTTTTC | 1009 |
| Qy | 844 | TAGTTTCAATTTATTCGTAGAGATGGTGACCGGCTATTTTTGTCTCAGAAACGAGTTGGAC | 903 |
| Db | 1010 | TAGTTTCACAATCAGAAAGATGGTGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC | 1069 |
| Qy | 904 | AGATAGGACGATATTTTACATTTACAGTTTCGATCGATGATGTTGTGATGCTGAGGAGC | 963 |
| Db | 1070 | GTAATGGTAGGATTTTATAGTTCTTATAAATTCAGATCAATGCGAGTAGTAGCAACAAA | 1129 |
| Qy | 964 | GCAAAAGACTTCTCAGCCAAACAGATGCAAGGTGGGTATGTTTAAATGGGAA | 1023 |
| Db | 1130 | TTAAGAAAGATTTATTTAGTTTCACAATCAAAATGACAGGG-----CTAATGTTTAAAGTTAGA | 1184 |
| Qy | 1024 | AAACGATCCTAGAAATTTACTCCAATTTGGACATTTTCATACGCAAAAACAAAGTTTACAGGAT | 1083 |
| Db | 1184 | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT | 1243 |
| Qy | 1185 | CGATGATCCTAGAAATTTACTAAAATAGGAAAATTTATTCG-AAAAACAAGCATAGATGAGT | 1243 |
| Db | 1243 | TACCAAGATTTTATATATGTTTTTAAATTTGGCGATATAGATCTAGTTGGTACACGTTCCACTTA | 1143 |
| Qy | 1244 | TGCCTCAATTTCTATATATGTTTTTAAAGGATGATATAGTTTGTAGTAGGAACACGCGCTCCCA | 1303 |
| Db | 1303 | CAGTTGATGAATTTTGAATAATATCTCTGCTCAAAAGACGATTTGAGTTTAAAAACAG | 1203 |
| Qy | 1304 | CAGTTGATGAATATGAAAAGTATATTAACGCGAAGACGACGCTTAGTTTTHAAGCAG | 1363 |
| Db | 1363 | GAATCACTGCTTTGTGCAAAATATCTGCTAGAAAATATATATCACTGATTTTGTGATGAATTCG | 1423 |
| Qy | 1423 | TTGCTTTGGAATTTGAGCATATTTGAGTATGATTTGGAATTTGGAATTTGGAATTTGGAATTTG | 1323 |
| Db | 1483 | TAAAGTTTAGATGTTCAATATATCAATGAATTTGCTATTTTGGTTCAGATATTAAGATTTATTC | 1483 |
| Qy | 1372 | TAAAGACAGTGAAGTTGTTATTTGTTGAGAGAGGGAAGTAAAGTAAAGTAAAGTAAAGTAAAG | 1543 |
| Db | 1543 | TAAGTAAATTTGTTTGGTTCCTTTCAGGGGGAATTTTGTACTCACTTGTATTTT | 1426 |
| Qy | 1544 | GAATATAATGAAAATTTGCTGCTGTTGCTTCAAGTGGTGGTCACTAGACACACTTTGAACCT | 1603 |
| Db | 1603 | GTTTAAACCGTTTTTGGAGGAAGAAACGTTTTTTTGGGTAAACATTTGATTAAGAGGATGC | 1486 |
| Qy | 1604 | TTTGAACCCATTTTGGGAAAAGAAAGATAGTTTGGGTAAACCTTTGATTAAGAGGATGC | 1663 |
| Db | 1546 | AAGAAGTCTTTTCAAGAAATGAAAATAATGATTCATGTTACTTTTCCCAACAAATCCCAATCT | 1723 |
| Qy | 1664 | TAGGAGTATTTCTAAGAGAGAGATTTGTTATATCATTTGCTCTTTCCCAACAAACCCGTAATGT | 1606 |
| Db | 1723 | CATTAAATTTAGTGAATAATTTCTTTAGCTTTTCAAAATTTTACGTGATGAGAAACCCAGA | 1783 |
| Qy | 1724 | CAAAACTTGGTAAATAATTTCTTCTAGCTTTTAAAGTCTTTAGAAAGAAAGACCCAGA | 1666 |
| Db | 1666 | TGTTATTTATCTGCTGCGCGCTGCTGCTCCCTCTTTTCAATCGGAAACCTATTT | 1843 |
| Qy | 1784 | TGTTATCATATCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT | 1726 |
| Db | 1726 | TGAGCAACAGCATTATATTCGAAGTATTTGATTCGAGTTTAAATAAATCTCATTTAACTG | 1903 |
| Qy | 1844 | TGGTTGTAAGACCGTTTATATAGAGTTTTCGACAGGATAGATTAACCACTTTTGACAGG | 1786 |
| Db | 1786 | AAAACTAGTTTATCCCGTAAACAGATTTTTTATTTGTTTTCAGTGGGAAAGAAATGAAAGGT | 1963 |
| Qy | 1904 | AAAAATTAGTCTATCTCTGTAACAGATTAATTTTATTTTTCAGTGGGAAAGAAATGAAAGGT | 1846 |
| Db | 1846 | ATATCTTAATCTATTAACCTTGGGAGTATTTTATTAATTTTATTAATTTTATTAATTTTATTA | 2023 |
| Qy | 1964 | TTATCTTAAGGCAATTTAAATTTAGGAGGATTTTTTAAATGATTTTGTACAGTGGGACA | 1906 |
| Db | 1906 | CATGAAACAGTTTAAATTCGATTTGAATAAAGAGATTCGATTTTATTTGAAAAAATTTGGAAGT | 2083 |
| Qy | 2024 | CATGAAACAGGCTTCAACCGTCTTTATTAAGAGAGTTTGAATTTTAAAGGACAGGTGCT | 1966 |
| Db | 1966 | ATAACCCAGCAATTTTATTTTCAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG | 2143 |
| Qy | 2084 | ATTGATCAAGAGTTTCAATTTCAACCGGTTACTCAGACTTTTGAACCTCAGAAATTTGTCAG | 2026 |
| Db | 2026 | TATAAAAAATTTCTCAGTTTCAAGAAATGGAACAATATATTAACAATCAGAAGTAGTT | 2203 |
| Qy | 2144 | TGGTCAAAATTTCTCTCATATGATGATGAATGAACTCTTTACATGAAGAGAGCTGAGATTTGT | 2086 |
| Db | 2086 | ATTTGCGCAGGAGCCCGCTACTTTTATGAATTTTATTTTCAATTTTCAATTTTCAATTTTCAAT | 2263 |
| Qy | 2204 | ATCACAATGCGGCTCAGCGACGTTTATGAATGACGTTTCTTAAGGGGAAAAAATCTATTT | 2146 |
| Db | 2146 | TTGTTTCTTAGACAAAAAAGTATGTTGAACATGTAATGATCATCAAGTAGAGTTTGTGA | 2323 |
| Qy | 2264 | GTGTTTCTTAGACAAAAAAGTATGTTGAACATGTAATGATCATCAAGTAGAGTTTGTGA | 2323 |

| | | | |
|------------|------|---------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| QY | 2147 | AGAGAAATTTTACAGATAATAATAATTTATTTATAGAAAAATATAGATGATTTGTTGAA | 2206 |
| Db | 2324 | AAAGAGTTATTTCTGAAATATGAATTAGATTATATTTTGAATATACAGTGAATTAGAGAA | 2383 |
| QY | 2207 | AAATTTATTGAAGTTCTTAAGCAAAAC--TAACTTTACATCAAAATAATAATTTTTTTTGT | 2263 |
| Db | 2384 | ATTATTTAGGAAAAAATAATATCTACTAGTAAAGTAATATACAAAACAATGATTTTGT | 2443 |
| QY | 2264 | GAAGATTAAAACAATAGTTGAAAAATTTTAATGAGGATCAAGAAAAATGAATATAAAAA | 2323 |
| Db | 2444 | TCCTCTTTCAAAAATGAACTTCTTAACTATTTTGAATAAATAATATTTTGTGGAGAAAA | 2503 |
| QY | 2324 | AGATGCA | 2330 |
| Db | 2504 | AATTGAA | 2510 |
| RESULT | 14 | | |
| AF363060 | | 4513 bp | DNA linear BCT 07-JAN-2002 |
| LOCUS | | | |
| DEFINITION | | | Streptococcus agalactiae strain SG99/077 capsular polysaccharide biosynthesis gene cluster, partial sequence. |
| ACCESSION | | | AF363060 |
| VERSION | | | AF363060.1 |
| KEYWORDS | | | GI:18029179 |
| SOURCE | | | |
| ORGANISM | | | Streptococcus agalactiae |
| | | | Streptococcus agalactiae |
| | | | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus. |
| REFERENCE | | | 1 (bases 1 to 4513) |
| AUTHORS | | | Kong, F., Gowan, S., Martin, D., James, G. and Gilbert, G.L. |
| TITLE | | | Serotype identification of group B streptococci by PCR and sequencing |
| JOURNAL | | | J. Clin. Microbiol. 40 (1), 216-226 (2002) |
| MEDLINE | | | 21635022 |
| PUBMED | | | 11773119 |
| REFERENCE | | | 2 (bases 1 to 4513) |
| AUTHORS | | | Kong, F. and Gilbert, G.L. |
| TITLE | | | Direct Submission |
| JOURNAL | | | Submitted (22-MAR-2001) CIDMLS, ICPMR, Westmead Hospital, Darcy Road, Sydney, NSW 2145, Australia |
| FEATURES | | | Location/Qualifiers |
| source | | | 1. .4513 |
| | | | /organism="Streptococcus agalactiae" |
| | | | /mol_type="genomic DNA" |
| | | | /strain="SG99/077" |
| | | | /serotype="Ia" |
| | | | /db_xref="taxon:1311" |
| gene | | | <1..126 |
| | | | /gene="cpsD" |
| CDS | | | <1..126 |
| | | | /gene="cpsD" |
| | | | /codon_start=1 |
| | | | /transl_table=11 |
| | | | /product="CpsD" |
| | | | /protein_id="AAL56395.1" |
| | | | /db_xref="GI:18029180" |
| | | | /translation="AKEQWESGSKFLGIILNKVSESVATYGDYGVYKDKRKK" |
| gene | | | 352..1527 |
| | | | /gene="cpsE" |
| CDS | | | 352..1527 |
| | | | /gene="cpsE" |
| | | | /codon_start=1 |
| | | | /transl_table=11 |
| | | | /product="CpsE" |
| | | | /protein_id="AAL56396.1" |
| | | | /db_xref="GI:18029181" |
| | | | /translation="WLVKYSFYIPISSLPFIKNSPFTTTLRSFSPFIAMNSILYL |
| | | | LNSFLKYRKYSYAKFSDTKVLITNKSLSKMTFRNKYDHYIAVCILDSKDCY |
| | | | DUKHSRLINKDAUTSELCTIVDQAPINIPLEFGYQIQDIINDIEMGVIVNN |
| | | | VEALGFNDIGERIGTIFEGYVTVYQMKFYKYIAKRFIDITGAILGILGIVAI |
| | | | FLVPIQRKDGPAIFSQNRVNGRIIFRFYFRSMRVDAEQIKDLIVHQMGLMFK |

293 TTTATCTTTCTGATTTTTCACAGACACTTTTGGAGTCGTGGCTACTCTTTGAAAGAGTTTAAAA 352
Qy
193 AAACATTTAACTAGTATATATTTTGAATTTTCTTACGGCAGTATCAATTTTGTGG 252
Db
353 TGGTATTTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATATTTTATTTTAT 412
Db
253 AGAATAATTTTGGCACTTTTCAAGACGTGGTCCGGTATTTTTCACATTAATAAATCTCGTTT 312
Qy
413 AAAACTCATTTTCAACAGCAGCACTTTTCCCTTTTTCCTTTTATTTTATTTGCTATGAAATTCGAAT 472
Db
313 TGGTATACCTATTTTAAAGTAAATTTTAAAGCAGTTTAAAGGATAGCTTTTCTATTTTTCGACAA 372
Qy
473 TATTGTATCTATTGAAATTCATTTTAAAAATATTTATCGAAAAATATTTCTTACGCTAAGTTT 532
Db
373 TCTATCAAAAAAGACAGATTTCTAATTTACACGGCTGAAACGATGGGAAAAATATGCAAGTTT 432
Qy
533 CACGAGATACCAAAAGTTTGTGTAACGAATAAGGATTTCTTTATCAAAAAATGACCTTTTA 592
Db
433 TATTTGAATCACATAAACAATAATTCAAAAAATCTTGTTCATTTGGTATGGTAGTTTAAAGTACAG 492
Qy
593 GGAATAAAT --- ACGACCATTAATTTATTTGCTGTCTGCATCTTGGACTCCTCTGAAAGG 649
Db
493 AAATAGATAAAATTTTATCATTTACCGCTCTATTATTTCTGTGGAAGAAAGCTATAGAGT 552
Qy
650 ATTGTTATGATTTGAAACATAAATCTGTTAAGGATAATAACAAGATGCTCTTACTTCAG 709
Db
553 TTTCAACAGGGAAGTGGTGCAGCACGCTTTTATAAATCTACAAAGTGAAGTTTTAGACG 612
Qy
710 AGTTAAGCTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCCATTTGAATTTTGGTA 769
Db
613 TAAAGCAATTC -----GTTTCAGATTTTGAAGTTTGTAGGTTATTTGATGAAGCGTTG 663
Qy
770 AATCCAAATACAAGATATTTAATGACATTTGAAGCAATGGGAGTGAATTTGTCATGTTA 829
Db
664 ATATTAATTTCACTCGGTTTTTACTCGGTTGAAAAACAAAAATCCAACTCGTAGGTGACC 723
Qy
830 ATGTAGAGGCATTTAGCTTTGATAATATAGGAGAAAGCGAATCCAACTTTTGAAGAT 889
Db
724 ATAGCATTTGAATTTTTCACAAATTTTATAAGCCTAGTCTATATCATGATGAAGACGAC 783
Qy
890 ATAGTGTATTACATATCTATGAAATTTCTATAAATATAGTACCTTTATAGCAAAACGAT 949
Db
784 TTTTGGATATACCTCGGAGCGTAGTCCGGTTAAATTTTCTGCGTATATTTCTATTTTGT 843
Qy
950 TTTTGGATATACCGGGTGTCTATTAATAGTTTGTCTATATGGAATTTGTGGCAATTTTTC 1009
Db
844 TAGTTCCAATTTATTCGTAGAGATGGTGCACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC 903
Qy
1010 TAGTTCACAAATCAGAAAAGATGTTGCAACCGGCTATCTTTCTCAAAATAGAGTAGGTC 1069
Db
904 AGAATGGAACGATATTTTACATTTTACAAAGTTTTCATCGATGATGTTGTGCTGAGGAGC 963
Qy
1070 GTAATGTTAGGATTTTATAGATTTCTATTAATTTTCAATCAATGCGAGTAGATGCGAACA 1129
Db
964 GCAAAAAGACTTCTCAGCCAAACACAGATGCAAGGTTGGGTATGTTTAAATGGAA 1023
Qy
1130 TTAAAGAAATTTTATAGTTTCAANTCAANTGACAGGG -----CTAATGTTTAAAGTTAGA 1184
Db
1024 AAACGATCTAGATTTACTCCAAATTTGACATTTTCTATAGCAAAACAAAGTTTAAAGCAGT 1083
Qy
1185 CGATGATCTAGATTTACTAAAATAGGAAATTTTATTCG -AAAAACAAGCATAGATGAGT 1243
Db
1084 TACCACAGTTTATAATGTTTAAATGCGGATATGAGTCTAGTTGGTATGACGCTCCACTA 1143
Qy
1244 TGCCTCAATTTCTATAATGTTTAAAGGTTGATATGAGTTTATAGGAAACACGCGCTCCCA 1303
Db
1144 CAGTTGATGATTTTGAATAATATCTCCTGCTCAAGAGACGATTTGAGTTTAAACCAG 1203
Qy
1304 CAGTTGATGATTTGAAGATTAATTTCAACGCAAGAGCGACGCTTAGTTTAAAGCCAG 1363
Db
1204 GGAATACAGGCTCTCGCAGGTTAGTGGTGGTATGATTAATATACAGACATTCGAGCAGCTAG 1263
Qy
1364 GAATCACTGGTTTGTGGCAAAATATCTCGTAGAAATATATCACTGATTTTGTGAAATCG 1423
Db

1264 TTCGGTTGGACTTAGCATACATTTGATTAATTTGGACTATCTCGTCAAGATTTAAAAATTTTAT 1323
Qy
1424 TAAAGTTAGATGTTTCAATATATCAATGAAATGGTCTATTTGGTCAAGATTTAAAGATTTATC 1483
Db
1324 TAAAGACAGTGAAGTTGTTATTTGAGAGAGGGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1372
Qy
1484 TCCTAACACTAAAGGTAGTCTTACTTTGGGACAGTGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1543
Db
1373 -----TATGAAAGTTTGTGTTGGTCCGTTCTTCCAGGGGACATTTTGACTCACATTTGTATTT 1426
Qy
1544 GAATATAATGAAATTTTGTCTGGTTCAAGTGGTGGTCACTAGCACACTTTGAACCT 1603
Db
1427 GTTAAACCGTTTGGAGGAAGAAACGTTTTTGGGTAAACATTTGATTAAGAGGATGC 1486
Qy
1604 TTTGAAACCCATTTGGGAAAAAGAGATAGTTTGGGTAAACCTTTTGTATTAAGAGATGC 1663
Db
1487 AAGAAGTCTTTTCAAGAAATGAAAAATGTATCCATTTACTTTTCCCAACAATCGCAATCT 1546
Qy
1664 TAGGAGTATTTCTAAGAGAGAGATTTGTATATCATTTGCTTCTTTCCAAACAACCCGTAATGT 1723
Db
1547 CATTAATTTAGTGAAAAATACTTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606
Qy
1724 CAAAAACTTGGTAAAAATACTATTTCTAGCTTTTAAAGTCTTTAGAAAAAGAAACGACAGA 1783
Db
1607 TGTATATTTTCACTGGTGGCGGCTTGTCTGCCCTTCTTTTATCATCGGAAACATTTT 1666
Qy
1784 TGTATATCATATCTGGTCCGCTGTAGCAGTACCAATTTCTTTTATATTTGTAAGTATTT 1843
Db
1667 TGCAGCAAGACGATTTATATTCGAAGTATTTGATTCGAGTTTAAATAATCTACATTTAACTGG 1726
Qy
1844 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACCACTTTTGAACAG 1903
Db
1727 AAAACTAGTTTATCCGCTAACAGATATTTTATTTGTTTCAAGTGGGAGAAAAATGAAGAAGT 1786
Qy
1904 AAAATTAGTGATTCCTGTAACAGATAAAATTTATTTGTTTCAAGTGGGAGAAAAATGAAGAAGT 1963
Db
1787 ATATCTTAATCTATTAACCTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAACT 1846
Qy
1964 TTAATCTAAGGCAATTTAATTTAGGAGGAATTTTAAATGATTTTGTACAGTGGGACCA 2023
Db
1847 CATGAACAACAGTTTAAATCGATTTGATAAAGAGATTTGATTTTGAAGAAAAATGGAAGT 1906
Qy
2024 CATGAACAGCGTTTCAACCGCTTTTAAAGAGAGTTGATAGATTTAAAGGGAACGAGTGCT 2083
Db
1907 ATAAACGACGAAATATTTTATTTCAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG 1966
Qy
2084 ATTGATCAAGAAGTGTTCATTTCAACCGGTTTACTCAGACTTTTGAACCTCAGAAATTTGCAG 2143
Db
1967 TATAAAAAATTTCTCAGTTTCAAGAAATGGAACAATATATTTAAACAATCAGAAAGTAGTT 2026
Qy
2144 TGGTCAAAATTTCTCTCATATGATGATTAATGAACTCTTTACATGAAGAAAGCTGAGATTTGT 2203
Db
2027 ATTTGCCACGGAGCCCGCTACTTTTATGAAATTTCAATTTCCAAAGGAAAAACAATTA 2086
Qy
2204 ATCACAATGCGCGGTCCAGCGAGCTTTATGAAATGCAAGTTTCTAAAGGGAAGAAACATTT 2263
Db
2087 TTGTTTCTTAGACAAAAAAGTA TGGTGAACATGTAATAATGATCATCAAGTAGAGTTTGTGA 2146
Qy
2264 GTGGTTCTTAGACAAAGACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGATTTTGTG 2323
Db
2147 AGAAGAAATTTTCAAGATATAATAATTTTATTTATGAAATAATATAGATTTTGTGTGAA 2206
Qy
2324 AAAGAGTTATTTCTTGAATAATGAAATTAGATTTATTTTGAATATCATGTAATTTAGGAAT 2383
Db
2207 AAAATTTTGAAGTTTCTAAGCAAC ---TAACTTTTACATCAAAATATAATTTTGTGT 2263
Qy
2384 ATTATTAAAGGAAAAAATATATCTATAGTAAAGTAAATATATCACAACAACATGATTTTGT 2443
Db
2264 GAAAGATTTAAAACAAAATAGTTGAAAAATTTTAAATGAGGATCAAGAAAAATGAATATAAAAA 2323
Qy
2444 TCCTCTTTCAAAATGAACTTTCTAAACTATTTTGAATAATAATATTTTGTGAGAAAAA 2503
Db

Db 482 TATTGTATCTATTGAAATTCATTTTAAAAATATATCGAAAAATATTTCTTAGCGTAAAGTTT 541
Qy 373 TCTATCAAAAAAGACAGATTTCTAATTACACGGCTGAACGATGGGAAATATGCAAGTTT 432
Db 542 CACGAGATACCAAAAGTTGTTTGTATACGAATAAGGATTTCTTTATCAAAAATGACCTTTA 601
Qy 433 TATTGTAATCACATAAACAANAATCAAAAAATCTTGTTGTCATTTGGTAGTTTGTAGTACAG 492
Db 602 GGAATAAT---ACGACCAATATTTATTTGCTGCTGCACTTGGACTCTCTGAAAGG 658
Qy 493 AAATPAGATAAAATTAATTTATCAATACCGCTCTAATTTCTGTGGAAGAAGCTATAGAGT 552
Db 659 ATTGTTATGATTTCAAAACATAAATCGTTTAAGGATAATAAACAAGATGCTCTTACTTCAG 718
Qy 553 TTTCACACAGGGAAGTGGTGCACACAGTCTTTTATAATCTACAAAGTGAGTTTGTAGACG 612
Db 719 AGTTAACTGCTTAACCTGTTGATCAAGCTTTTATTAACAATACCAATTCATTTTGGTA 778
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTGAGTTGTTAGGTAATGATGTAAGCGTTG 663
Db 779 AATACCAAAATACAGATATTTAATGACATGAGCAATGGGAGTGATTTGTCATGTTA 838
Qy 664 ATATTAATTCATTCGGTTTACTCGGTTGAAAAAACAANAATCCAACTGCTAGGTGACC 723
Db 839 ATGTAGAGGCATTTAGCTTTGATAATATAGGAGAAAGCGAATCCAACTTTTGAAGAT 898
Qy 724 ATAGCATTTAACTTTTCCAAAATTTTATAAGCCTAGTCATATCATGATGAAGACGAC 783
Db 899 ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT 958
Qy 784 TTTTGGATATACTCGGAGCGGTACTCGGGTTAAATTTATTTGTTGATAGTTTCTTATTTGT 843
Db 959 TTTTGGATATCACGGGTGCTAATTAATAGTTTGTCTATATGTCATATGGCAATTTTTC 1018
Qy 844 TAGTTCCAAATTAATTCGTAGAGATGGTGCACCGGCTAATTTTGTCTCAGAAAACGAGTTGGAC 903
Db 1019 TAGTTCCACAAATCAGAAAAGATGGTGCACCGGCTAATTTTCTCAAAATAGAGTAGTGC 1078
Qy 904 AGAATGAGCGATATTTACATTTACAAAGTTTCGATCGATGATGTTGATGCTGAGGAGC 963
Db 1079 GTAATGTTAGGATTTTGTAGATTTCTATAAATTCAGATCAATGCGAGTAGATGCGAACAA 1138
Qy 964 GCAAAAAGACTTCTCAGCCAAACAGATGCAAGGTGGGTATGTTTAAATGGAA 1023
Db 1139 TTAAGAAGATTTATTTAGTTTCAATCAATATGACAGG-----CTAATGTTAAGTTAGA 1193
Qy 1024 AAACGATCCTAGAAATTTACTCCAATTTGACATTTTCATACGCAAAAACAAAGTTTAGACGAGT 1083
Db 1194 CGATGATCCTAGAAATTTACTAAAATAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT 1252
Qy 1084 TACCACAGTTTATAATGTTTAAATTTGCGGATATGAGTCTAGTTGGTACAGCTCCACTA 1143
Db 1253 TGCCTCAATTTCTAATGTTTAAAGGTGATATGAGTTTAGTAGGAACAACGCGCTCCCA 1312
Qy 1144 CAGTTGATGATTTGAAAAATATACTCTGCTCAAGAGACGATTCAGTTTAAACAG 1203
Db 1313 CAGTTGATGATATGAAAAGTATAATTTCAACGCAAGACGCGCTTAGTTTTAAGCCAG 1372
Qy 1204 GGATTACAGGTTCTCGGCAAGTTAGTGTCTGATGATATACACAGATTCGACGACGTAG 1263
Db 1373 GAATCATTGTTTGGCAATATCTGGTAGAATATATACATGATTTTGTAGAAATCG 1432
Qy 1264 TTCGGTTGGACTTAGCATATGATTAATTTGGAATCTATCGTTCAGATATTAATAATTTAT 1323
Db 1433 TAAAGTTTAGATGTTCAATATATCAATGAATGCTTATTTTGGTCAGATTAAGATTTATC 1492
Qy 1324 TAAAGACAGTCAAAAGTTGTTATTTGTCAGAGGGAAGTAAAGTA----- 1372
Db 1493 TCCTAAACACTAAAGGTAGTCTTACTTGGGACAGGTGCTAAGTAAAGGTAAAGTTTGAAG 1552
Qy 1373 -----TATGAAAGTTGTTTGGTTCGTTCTTACAGGGGACATTTGACTACACTGTTATTT 1426
Db 1553 GAATATAATGAAATTTGCTGTTGTTCAAGTGGTGTCTATCTAGCACACTTTGAACCT 1612

Qy 1427 GTTAAACCGTTTGTGAAGGAAGAAACGTTTTCGGTAAACATTTTGATAAAGAGGATGC 1486
Db 1613 TTTGAAACCCATTTGGGAAAAAGAGATAGGTTTGGTAAACCTTTTGATAAAGAGATGC 1672
Qy 1487 AAGAAGTCTTTTGAAGAAATGAAAAATGTATCCATGTTACTTTCCAAACAATCGCAATCT 1546
Db 1673 TAGGAGTATTTCTAAGAGAAGAGATTTGTATATCATGCTCTTTCCAAACAACCGTAATGT 1732
Qy 1547 CATTAATTTAGTCAAAAAATCTTTCTTAGCTTTCAAAATTTTACGTTAGTATGAGAAACGAGA 1606
Db 1733 CAAAAATCTGGTAAAAAATACTATTCTAGCTTTTAAAGTCTTTAGAAAAAGAAACGACAGA 1792
Qy 1607 TGTATTATTTTCACTGTTGGCGCGTTCGTCCTCCCTTCTTTTACATCGGAAACATTT 1666
Db 1793 TGTATCATATCATCTGGTCCGCTGTAGCAGTACCAATCTTTTATATATTTGTAAGTTATT 1852
Qy 1667 TGAGCAAAAGACCATTTATATGAAGTATTTGATCGAGTTAAATAATCTACATTAACCTGG 1726
Db 1853 TGGTTGTAGACCGTTTATATAGAGTTTTCGACAGATAGATAAACCACTTTTGACAGG 1912
Qy 1727 AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCACTGGGAGAAATGAAGAAGT 1786
Db 1913 AAAATTAGTGTATCTCTTAACAGATAAATTTATGTTTCACTGGGAGAAATGAAAAAGT 1972
Qy 1787 ATATCTTAATCTATTAACCTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAT 1846
Db 1973 TTATCTCAAGGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTTCACAGTGGGACA 2032
Qy 1847 CATGAACAACAGTTTAAATCGATTGATAAAGAGATTTGATTTATTTGAAAAAATGGAAGT 1906
Db 2033 CATGAACAGCGTTTCAACCGTCTTTTAAAGAGTTGATAGATTTAAAGGACAGGTGCT 2092
Qy 1907 ATAACCGACGAATATTTATTTCAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG 1966
Db 2093 ATTGATCAAGAAGTGTTCATTCAACCGGTTACTTCAGACTTTTGAACTTCAGAAATTTGTCAG 2152
Qy 1967 TATAAAAAATTTCTCAGTTTCAAGAAATGGAACAATATATTTAAACAATCAGAAGTAGTT 2026
Db 2153 TGGTCAAAATTTCTCTCATATGATGATGAACTCTTACATGAAGAAGCTGAGATTGTT 2212
Qy 2027 ATTTGCCACGAGGCCCGCTACTTTTATGAAATTCATTATCCAAAGCAAAAAACAATTA 2086
Db 2213 ATCACAATGCGCGTCCAGCGAGTTTATGAAATGCGAGTTTCTAAAGGGAANAACATATT 2272
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGTGA 2146
Db 2273 GTGGTTCTTAGACAAGAACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGATTTTGTG 2332
Qy 2147 AGAAGATTTTACAGATAATAATATTTTATTTATAGAAAATATAGATGATTTGTTGAA 2206
Db 2333 AAAAGAGTTATTCTTGAATAATGAATAGATTATTTTGAATATCAGTGAATTAGAGAA 2392
Qy 2207 AAAATTTTCAAGCTTTCTAAGCAAC---TAACTTTACATCAAAATAATAATTTTGTG 2263
Db 2393 ATTATTAAAGAAAAAATATATCTACTAGTAAAGTATATACAAAACAATGATTTTGT 2452
Qy 2264 GAAAGATTTAAACCAATAGTTGAAAAATTTTAAATGAGGATCAAGAAAATGAATATAAATA 2323
Db 2453 TCCTCTTTCAAAAAATGAACCTTCTAAACTATTGATAAATAATATTTTGTGGAGAAAAA 2512
Qy 2324 AGATGCA 2330
Db 2513 AATTGAA 2519

RESULT 16
AF332893
LOCUS
DEFINITION
Streptococcus agalactiae strain 090 capsular polysaccharide
biosynthesis gene cluster, partial sequence.
AF332893
AF332893
VERSION
AF332893.1 GI:18033311

KEYWORDS

SOURCE ORGANISM

Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 4888)
Kong, F., Gowan, S., Martin, D., James, G. and Gilbert, G.L.
Serotype identification of group B streptococci by PCR and sequencing
J. Clin. Microbiol. 40 (1), 216-226 (2002)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. 4888
/organism="Streptococcus agalactiae"
/mol_type="genomic DNA"
/strain="090"
/serotype="Ia"
/db_xref="taxon:1311"
<1. 135
/gene="cpsIaD"
<1. 135
/gene="cpsIaD"
/codon_start=1
/transl_table=11
/product="CpsIaD"
/protein_id="AAL57052.1"
/db_xref="GI:18033312"
/translation="AKEQMEQSGSKFLGIILNKVSESVATYGDYDGYNGYKDRKREK"

gene

CDS

361. 1536
/gene="cpsIaE"
361. 1536
/gene="cpsIaE"
/codon_start=1
/transl_table=11
/product="CpsIaE"
/protein_id="AAL57053.1"
/db_xref="GI:18033313"
/translation="MWLKSPYIFIFISSLPFIKNSFTTTLRSLFFPSFIAMNSILLVL
LNSFLKYRYKYSAKFSRDTKVLIITKDSLSKMTFRKYNHVIANCILDSSSEKCY
DLKNSRLINKDALTSELCTCLTDQAFINIPILFKYQIQDIINDIEAMGVIVYN
VEALSFONIGKRIQTEGVSVIITYSMKPKYSHLIAKRFIDITGAILGLLIGIVAI
FLVPIQRDGPAPAFSONVRNGRIIFRFYKFSMRVDAEQIKDILLVHQMOTGLMFI
LDDPRITKTKGKFRKTSIDELPOFYNVLKGDMSLVGTRPTVDVEYKYNSTQKRRLS
FKPGITGLWQISGRNNITDPEIIVKLDVQYINWSIDIKIILLTLKVLVLTGAK"
1560. 2009
/gene="cpsIaF"
1560. 2009
/gene="cpsIaF"
/codon_start=1
/transl_table=11
/product="CpsIaF"
/protein_id="AAL57054.1"
/db_xref="GI:18033314"
/translation="MKICLVSGSGHLAHLNLLKPIWEKEDRFVWTFDKEDARSILRE
EIVVHCFFPTNRNKNLIVKNTILAFKVLKRPDVIISGGAIVAPRFYIKLPGCKT
VYIEVFDRIKPTLTKLVIPVTDKFIQVQWEEKVYPKAINLGGIF"
2009. 2491
/gene="cpsIaG"
2009. 2491
/gene="cpsIaG"
/codon_start=1
/transl_table=11
/product="CpsIaG"
/protein_id="AAL57055.1"
/db_xref="GI:18033315"
/translation="WIFVTVGTHEQQFNRLIKEVDRLKGTGAIQDEVFIQGYSDPEP"

QNCQWSKFLSYDDMNSNMKEABIVITHGGPATFNNVSKGKKTIVWRQEQFGHVN
HQVDFLKLFLKYELDIYILISELENIKEKNISTSVKISQNDPFCSPKNSKLFE
" 2587. 3726
/gene="cpsIaH"
2587. 3726
/gene="cpsIaH"
/codon_start=1
/transl_table=11
/product="CpsIaH"
/protein_id="AAL57056.1"
/db_xref="GI:18033316"
/translation="MQGLLLALIVLLICSSYNEKMKFLNMBEFFVIFYIYLTSL
LHSLFKNPDRILANFSNLTIGIVSLAFHYKNTTLEDKILKAFNLGLLIFFLG
GTYCYLHNINQISIFGRDLIGDWMNGHMTQAMGFFYSNLIIPITVVTNYIYI
YMLRYSIMTVIGVLLFTPLIPIGSGRAGIVAILAQMTILLANTVVKKTKFGL
YILPFLAVIMVLYFDNLSTIYRIILRSGSSSRFSLYKDTNIVNINLLGKGV
KELWNSDLPLGSHSTYIGFYKGLGLMNVPEGLLLIFNIGRKAQKQSAFYBEIG
TLITLFSFALDEDLGGANWLIVFTVLGILENKKDFYSQLKRWKS"
3726. 4700
/gene="cpsIaI"
3726. 4700
/gene="cpsIaI"
/codon_start=1
/transl_table=11
/product="CpsIaI"
/protein_id="AAL57057.1"
/db_xref="GI:18033317"
/translation="MEKRLIVSIIPIYNSEAYLKECVQSVLQOHPLEIIVLLIDGS
TDNSGEICDNLSDQEDNRILFHRKNGGVSSARNLGDKSTGEFITFVDSDDFVAPNMI
EIMLKNLTINADIAEVDIFDSNERDYKRRKRNRYKVFKNNSKFLSGNRVNI
CTKYKSIIGNLRFDENLIGEDLFCNKLCOEHRIVVDVTSLSYTYRIVKTSAMN
QKFNESLDITTFINEVSSLPVAKLANVEAKPLREKIKLKRKMFELSGNIDNKIKVQ
RELFPKDKSVYPPFYKAVKYLKLLSFYLMKCSPLKLVVMAYRFRQK"
4734. 4888
/gene="cpsIaJ"
4734. 4888
/gene="cpsIaJ"
/codon_start=1
/transl_table=11
/product="CpsIaJ"
/protein_id="AAL57058.1"
/db_xref="GI:18033318"
/translation="MIKKIEKDLISVIVPIYVNDVYLVECTESLIVOTYRNIILLIN
DGSNDNCA"
ORIGIN
Query Match 10.4%; Score 729.4; DB 1; Length 4888;
Best Local Similarity 59.2%; Pred. No. 2.6e-73;
Matches 1390; Conservative 0; Mismatches 916; Indels 41; Gaps 7;
Qy 16 TGGCAATTATTTGATATAGTACAGTTGCAATTTCTGCAATCTTAAACAGTCATATACCAA 75
Db 182 TAGCGATGATTCAACAGTTGTGGTTATTTTCTGCAAGTTTGACATTAACATTAATTA 241
Qy 76 ATGCTGATTTAAATCG--TTCTGGAATTTTATCAATCATGATGTTTCATTATTTTGCAT 132
Db 242 CTCCCAATTTTAAAGCAATAAGATTATTTGTTTCTTATGATACATTATATTTGCT 301
Qy 133 TTTTATATCTCGTATCCAGTTTGAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAAA 192
Db 302 TTTATCTTCTGATTTTACAGACACTTTTGGAGTCGTGGCTACTCTTGAAGAGTTTAAAA 361
Qy 193 AAACATTAACTATAGTATATATTTTGCATTTTCTTACGGAGTATCATTTTGTGTTGG 252
Db 362 TGGTATTTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTATTTTATA 421
Qy 253 AGAATAATTTCCGCACTTTCAAGCGTGTGCGGTATTTTCCACATTAATTAACCTCGTTT 312
Db 422 AAAACTCATTTACAAOCACAGCACTTTCCTTTTTTCTTTTATTTGCTATGAATTCGATT 481
Qy 313 TGGTATACCTATTAAACGTAATTAATTAAGCAGAGTTTAAAGGATAGCTTTCTATTTTCGACAA 372

Db 482 TATTGTATCTATTGAAATTCATTTTAAATAATATATCGAAATAATCTTTACGCTAAGTTTT 541
Qy 373 TCTATCAAAAAGACAGATCTCTAATTAACAAGCGCTGAACGATGGGAAAAATATGCAAGTTTT 432
Db 542 CACGAGATACCAAGTGTGTTGATAACGAATAGGATCTTTATCAAAAAATGACCTTTA 601
Qy 433 TATTGTAATCACATATAACAATTCAAAAAATCTTTGTCATCTGGTAGTTTGGTAGAC 492
Db 602 GGAATAAAT---ACGACATAATATATATGCTGTCGATCTCTGGAACCTCTGTAAGG 658
Qy 493 AAATAGATAAAATTAATTTATCATTTACCGCTCTATTTATCTGTGGAAGAAGCTATAGAT 552
Db 659 ATTGTTATGTTGAACATAAATCTGTTAAGGATAAACAAGAATGCTCTTACTTTCAG 718
Qy 553 TTTCAACAAGGAAGTGGTGCACACGCTCTTTATAAATCTACCAAGTGAATTTTAGACG 612
Db 719 AGTTAACTGCTTAATCTGTTGATCAAGCTTTTATTAACATACCCATTTGAATTTTGGTA 778
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTGAGTTGTTAGGTATTTGATGTAAGCGTTG 663
Db 779 AATACCAATACAAAGATTAATTAATGACATTTGAAGCAATGGGAGTGAATTTGCAATGTTA 838
Qy 664 ATATTAATTCATTCGGTTTTACTCGGTTGAAAAACAAAAAATCCAACTGCTAGGTGACC 723
Db 839 ATGTAGAGGCACTTAGCTTTGATAATATAGAGAAAGCGAATCCAACTTTTGAAGGAT 898
Qy 724 ATAGCATGTAACTTTTCCAAATTTTATAAGCCTAGTCATATCATGATGAACGAC 783
Db 899 ATAGTGTATTACATATCTATGAAATCTTATAAATATATGTCACCTTATAGCAAAACGAT 958
Qy 784 TTTTGGATATACTCGGAGCGGTAGTCGGGTTAATTTATTTGTTGGTATATGTTCTTATTTGT 843
Db 959 TTTTGGATATACGGGGTGTATTATAGGTTTGTCTATATGGCAATTTGGCAATTTTTC 1018
Qy 844 TAGTTCCAATTAATTCGTAGAGATGGTGCACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC 903
Db 1019 TAGTTCACAAATCAGAAAGATGGTGGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC 1078
Qy 904 AGAATGACGCATATTACATTTCAAGTTTTCGATCGATGTAATGTTGATGCTGAGGAGC 963
Db 1079 GTAATGCTAGGATTTTATAGTTCTATATAATTCAGATCAATGCGAGTAGATGCGAAACAA 1138
Qy 964 GCAAAAAGACTTCTCAGCCCAACCAACGATGCAAGGTTGGGTATGTTTAAATGGGAA 1023
Db 1139 TTAAGAAAGATTTTATAGTTTCAATCAATCAATGACAGGG-----CTAATGTTAAGTTAGA 1193
Qy 1024 AAACGATCTAGAAATTAATCTCAATTTGACATTTTCATACGCAAAAACAAAGTTTACAGAGT 1083
Db 1194 CGATGATCTTAGAATTAATAATAGGAAATTTATTTTCG-AAAAACAAAGCATAGATGAGT 1252
Qy 1084 TACCACAGTTTATAATGTTTAAATTCGCGATATGAGTCTAGTTGGTACACGTCACCTTA 1143
Db 1253 TGCCTCAATTTCTAATAATGTTTAAAGGTGATATGAGTTTAGTAGAAACACGCGCTCCCA 1312
Qy 1144 CAGTTGATGAATTTGAAAATATACCTCTGCTCAAAAGACGATTCAGTTTAAACCAG 1203
Db 1313 CAGTTGATGAATTAAGAAAGTATTAATTTCAACGCAAGAGCGACGCTTAGTTTAAAGCCAG 1372
Qy 1204 GGATTAACAGTCTCTGCGAGGTTAGTGGTGTAGTAAATATATACAGACTTCGACGACGTAG 1263
Db 1373 GAATCACTGGTTTGTGGCAATATCTGTAGAAATATATATCATGATTTTGAAGAAATCG 1432
Qy 1264 TTCGGTTGGACTTAGCATATCAATTTGATTAATTTGGACTATCTGGTCAGATATTTAAATTTAT 1323
Db 1433 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTAAGATTATTC 1492
Qy 1324 TAAAGACAGTGAAGTTGATTTGTTGAGAGGGAAGTAAGTAAGTA----- 1372
Db 1493 TCCTAACTAAAGGTAGTCTTACTTGGGACAGGTCTAAGTAAAGTAGGTTTGAAG 1552
Qy 1373 -----TATGAAAGTTTGTGTTGGTCTTCTCAGGGGGACATTTGACTCACTTGTATTT 1426
Db 1553 GAATATAATGAAATTTGTTCTGGTTGTTTCAAGTGGTGGTCACTCTAGCACACTTGAACCT 1612

RESULT 17
AF363058
LOCUS
DEFINITION
Streptococcus agalactiae strain Fp6453 capsular polysaccharide
biosynthesis gene cluster, partial sequence.
AF363058
AF363058.1 GI:18029163
VERSION

Qy 1427 GTTAAAAACCGTTTGGAGGAAGAGAAACGTTTTTGGGTAAACATTTGATTAAGAGGATGC 1486
Db 1613 TTTGAAACCCATTTGGGAAAAAAGAGATAGGTTTGGGTAAACCTTTGATTAAGAGGATGC 1672
Qy 1487 AGCAAGTCTTTTGAAGAATGAAAAATGATCCATGTTACTTTTCCCAACAATCGCAATCT 1546
Db 1673 TAGGAGTATTTCTAAGAGAAGAGATTGTAATATCATGCTTCTTCCCAACAACCGTAATGT 1732
Qy 1547 CATTAATTTAGTGAATAAATACCTTTCTAGCTTTCAAAAATTTTACGTGATGAGAAACGAGA 1606
Db 1733 CAAAACTTGGTAAAAAATACATTTCTAGCTTTTAAGGTCCTTAGAAAAAGAACCCAGA 1792
Qy 1607 TGTATTAATTTCACTGGTGGCGCGTGTGTCCTCTCTTTTACATCGGAAACATTT 1666
Db 1793 TGTATCATATCATCTGGTGGCGCTGTAGCAGTACCATCTTTTATATTTGTAAGTTATTT 1852
Qy 1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTTAATAATCTCATTTAACTGG 1726
Db 1853 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGATAGATAAAACCAACTTTTGACAG 1912
Qy 1727 AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAGTGGGAGAAAATGAAGAAGT 1786
Db 1913 AAAATTTAGTGTATCTCTGTAACAGATAAATTTATTTGTTTCAGTGGGAGAAAATGAAGAAGT 1972
Qy 1787 ATATCCTAAATCTATTAATCTTGGGAGTATTTTTTAATGATTTTGTAAACAGTAGGAAT 1846
Db 1973 TTATCCTAAGGCAATTTAAATTTAGGAGGAAATTTTAAATGATTTTGTACAGTGGGAGCA 2032
Qy 1847 CATGAACAACAGTTTAAATCGATTTGAAGAGATTTGATTTTATGAAAAAATTTGGAAGT 1906
Db 2033 CATGAACAGCAGTTTCAACCGTCTTATTAAGAAAGTTGATAGATTAAAGGAGCAGGTGCT 2092
Qy 1907 ATAAACGAGCAATAATTTTATTTCAACAGGATATTTCTGACTATATTTCCAGATAATTTGCAAG 1966
Db 2093 ATTGATCAAGAGTGTTCATTCAACCGGTTACTCAGACTTTGAACTCTCAGAAATTTGTCAG 2152
Qy 1967 TATAAAAAATTTCTCAGTTCAAAAGAAATGGAAACAATATATTAACAATCAGAAATGAGTT 2026
Db 2153 TGGTCAAAAATTTCTCTCATATGATATGAACTCTTTACATGAAAGAAAGCTGAGATTGTT 2212
Qy 2027 ATTTCGACGAGGCGCCGCTACTTTTGAATTTTCAATTTATCCAAAGGAAAAACAATTA 2086
Db 2213 ATCACACATGGCGGTCCAGCGAGTTTATGAATGCGAGTTTCTAAAGGGAAGAAAACTATT 2272
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGCTGAACATGTAATCATCATCAAGTAGAGTTTGTGA 2146
Db 2273 GTGGTTCTTAGACAAAGACAGTTTGGAGAGCATGTGATAATCATCAGGTGGATTTTGTG 2332
Qy 2147 AGAAGAAATTTTACAAGATAATAATATTTTTATATAGAAATAATAGATGATTTGTTGAA 2206
Db 2333 AAAGAGTTATTTCTTGAATAATAGAAATTTAGATTATTTTGAATATCATCAGTGAATTAGAAT 2392
Qy 2207 AAAATTTTGAAGTTTCTAAGCAAC---TAACTTTTACATCAAAATTAATATTTTGTG 2263
Db 2393 ATTATTAGGAAAAAATAATATCTCTAGTAAAGTAAATATCACAAAACAATGATTTTGT 2452
Qy 2264 GAAAGATTAAAACAAAATAGTTGAAAAATTTAAATGAGGATCAAGAAATTAATTAATAAAA 2323
Db 2453 TCCTCTTTCAAAAATGACTTTCTTAACTATTTTGAATAATAATATTTTGTGGAGAAAA 2512
Qy 2324 AGATGCA 2330
Db 2513 AATTGAA 2519

KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 4888)
REFERENCE Kong, F., Gowan, S., Martin, D., James, G. and Gilbert, G.L.
AUTHORS Serotype identification of group B streptococci by PCR and
TITLE sequencing
J. Clin. Microbiol. 40 (1), 216-226 (2002)
JOURNAL
MEDLINE 21635022
PUBMED 11773119
REFERENCE 2 (bases 1 to 4888)
AUTHORS Kong, F. and Gilbert, G.L.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2001) CIDMLS, ICPMR, Westmead Hospital, Darcy
Road, Sydney, NSW 2145, Australia
FEATURES
source
location/Qualifiers
1..4888
/organism="Streptococcus agalactiae"
/mol_type="genomic DNA"
/strain="FP6453"
/serotype="1a"
/db_xref="taxon:1311"
1..135
/gene="cpsD"
1..135
/gene="cpsD"
/codon_start=1
/transl_table=11
/product="CpsD"
/protein_id="AAL56381.1"
/db_xref="GI:18029164"
/translation="AKEQEQSGSKFLGILNKNVSVATYGDYGDYGYKXDRKRK"
"
361..1536
/gene="cpsE"
361..1536
/gene="cpsE"
/codon_start=1
/transl_table=11
/product="CpsE"
/protein_id="AAL56382.1"
/db_xref="GI:18029165"
/translation="WVLKYSEYVIFLSSSLFELFKNSFTTTLSPSFAMNSLLYL
LNSPLKRYKYSVAKRSRTKVLIITKDSLSKMTFRNKYQIDNYIAVCILDSKDCY
DKHNSLRILNKDALSELCTVDAQFINIPLEFGKYQIDINDIANGVIVNVN
VVALSPDITGEKRIQTFEGYSVITYGYKFKYSHLAKRFLDITGAILGLICGIVAI
FLVQIRKDGGAIPFQNRVGRNGRIFRFYKFRSMRVDAEQIKDLVHNMOTGLMFK
LDDPRITKIGKFIKTSIDELPQFYNVLKGDMSLVGTRPTVDEYKYNSTQKRLS
FRPGITGLQISGRNNITDFDEIVKLDVQYINWSIWSDKIILLTLKVVLLGTGAK"
1560..2009
/gene="cpsF"
1560..2009
/gene="cpsF"
/codon_start=1
/transl_table=11
/product="CpsF"
/protein_id="AAL56383.1"
/db_xref="GI:18029166"
/translation="MKICLVGSGGHLAHLNLLKPIWEKEDREFWTFDKEDARSILRE
EIVYHCFPTNRNVKNLVKNTILAPKVLKRPDVIISGAAVAVPFFYIGKLGCGKT
VILEVDRIDKPLTKLVYPTDKFIVQWEEMKVPYKPAINLGGIF"
2009..2491
/gene="cpsG"
2009..2491
/gene="cpsG"
/codon_start=1
/transl_table=11
/product="CpsG"
/protein_id="AAL56384.1"
/db_xref="GI:18029167"
/translation="MIFVTVGTHEQQFNRLIKEVDRLKGTGAIDQEVFIQTGYSDPEP"

QNCOWSKFLSYDDMNSYMKBAEIVITHGGPATFMNAVSGKKTIVVPRQEQGEHVN
HQVDFLKELFLKELDYILINSELENI IKERNISTSKVISQNNDFCSFKNLSKLF
"
2587..3726
/gene="cpsH"
2587..3726
/gene="cpsH"
/codon_start=1
/transl_table=11
/product="CpsH"
/protein_id="AAL56385.1"
/db_xref="GI:18029168"
/translation="MQLLLALLIVLLICSSNEMKFLNMAEIFFIVFIIVLTSL
LHSLFKNPDNRIILAAFNLSIIGIVSLAFHYHYKNTTLELDKILKAFNLGLLFLFG
GTYVCLHNINQNIIFGRDLIGSDWINGMHTQRAMGFYSNLIPIITVVTNLYIYI
YMLKNSYIMTIGVLLFTFPIPGSGRAGIVAILAOMFILLINTVVKKTKIFLL
YELPALLVMMLYFDNLLSIYIRINLRSSSESSEFSLYKDTNIVINNLSLFGGV
KELWNSDLPISGHSYTYGYFKYSGLLGLMNVPGLLIFTNIGRAKQSAFYEIVG
TLTILFSFPALEDLDGANLVIIFITVLGILENKKDFYQLKRWKS"
3726..4700
/gene="cpsI"
3726..4700
/gene="cpsI"
/codon_start=1
/transl_table=11
/product="CpsI"
/protein_id="AAL56386.1"
/db_xref="GI:18029169"
/translation="MEKRILVSIPIYNSEAYLKECVQSLQOOTHLEVLIDDGS
TDSGEICDNLQEDNRILVFKKGGVSSARNGLDKSTGEFTFVDSDDFVAPNMI
EIMLKNLITENADIAEVDIDISNERDYKRRNFYKFNNSNLSKEFLSGNKNVNI
CTKLYKKSIIIGNRIDEENLKI GEDLLFNCKLCOEHRIYVDTTSSLYTRI VKTSAMN
QKFNESLDFTITFNEVSLVPAKLANYVEAKFLREKIKLARMFELGNSINDNKIKVQ
REIFFDKIKSYFPYKAVKYLKLLSLFVLMKCSPLKYVMAYRRFQK"
4734..>4888
/gene="cpsJ"
4734..>4888
/gene="cpsJ"
/codon_start=1
/transl_table=11
/product="CpsJ"
/protein_id="AAL56387.1"
/db_xref="GI:18029170"
/translation="MIKIEKDLISVIVPYNVEDYLVCEIESLIVQTYRNIIEILLIN
DGSITDCA"
ORIGIN
Query Match 10.4%; Score 729.4; DB 1; Length 4888;
Best Local Similarity 59.2%; Pred. No. 2.6e-73;
Matches 1390; Conservative 0; Mismatches 916; Indels 41; Gaps 7;
QY 16 TGGCATATTATGATAGACAGTTGCAATTCCTGCAATCTTAAACAGTCATATACCA 75
DB 182 TAGCGATGATCAACAGTTGTGGTTTATTTTCTGCAAGTTTGACATTAACATTA 241
QY 76 ATGCTCAATTAATCG---TTCGGAAATTTTATCATATGATGATGTTTCAATTTTGCAT 132
DB 242 CTCCTCAATTTTAAAGCAATTAAGATTTATTTGTTCTTATTGATACATATATGCT 301
QY 133 TTTTATATCTCGTATGCCAGTTGAATTTAGATPATAGAGTAATCTGATAGAGTTGAA 192
DB 302 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAA 361
QY 193 AAACATTTAATATAGTATAATATTTGCAATTTTCTTACGGCAGTATCATTTTGTGG 252
DB 362 TGGTATTGAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTATTT 421
QY 253 AGAATAATTTCCGACTTTTCAAGACGGTGGCGGTGATTTTACATTAATAAATCTGTT 312
DB 422 AAAACTCATTTTAAACGACACGACTTTCCTTTTCTTTTATTTGATGAAATTCGATT 481
QY 313 TGGTATACCTATTTAACGTAATTTTAAAGCAGTTTAAAGGATAGCTTTCTATTTTCGAA 372

Db 482 TATTGTATCTATTGAATTCATTTTTTAAATAATTTATCGAAAAATATCTTACGCTAAGTTTT 541
QY 373 TCTATCAAAAAGACGATCTTAATTTACAACGGCTGAACGATCGGAAAAATATGCAAGTTTT 432
Db 542 CACGATACCAAGTGTGTTTGTATACGAATAAGGATCTTTATCAAAAAATGACCTTTTA 601
QY 433 TATTTCAATCACATAAACAATAAATCAAAAAATCTCTGTGCATTTGGTGTAGTTTGTAGGTACAG 492
Db 602 GGNATAAAT---ACGACCAATAATATATTTGCTGTCTGCAATCTTGGACTCTCTGAAAAGG 658
QY 493 AAATAGATAAAATTAATTTTATCATTAACCGCTCTATTAATTTCTGTGGAAGAGCTATAGAGT 552
Db 659 ATTGTTATGATTTGAAACATAACTCGTTAAGGATATAAACAAGATGCTCTTACTTCAG 718
QY 553 TTTCAACAAGGAGTGTGCGACCGCTCTTTATAAATCTACAAGTGTAGTTTTGTAGAGG 612
Db 719 AGTTAACTGCTTAACTGTGTGATCAAGCTTTTATTAACATACCCATTGAATTAATTTGGTA 778
QY 613 TAAAGCAATTC-----GTTTCAGATTTTGAGTTGTGTAGTATTGATGTAAAGGTTG 663
Db 779 AATACCAATAACAGATATTAATATGACATTAAGCAATGGAGTGTATGTCATATGTTA 838
QY 664 ATATTAATTCATTCGGTTTTTACTGCGTTGAAAAACAAAAAATCCAACTGCTAGGTGACC 723
Db 839 ATGTAGAGGCACCTAGCTTTTGATAATATAGGAGAAAGCGAATCCAACTTTTGAAGGAT 898
QY 724 ATAGCATTTGAATTTTTCACAAATTTTATAAGCTTAGCTAGTCAATATCATGATGAACGAC 783
Db 899 ATAGTGTATACATATCTATGAAATCTATAAATATAGTCACTTATAGCAAAACGAT 958
QY 784 TTTTGGATATCTCGAGCGGTAGTCGGGTAAATTTTGTGTGTAGTATTTCTATTTTGT 843
Db 959 TTTTGGATATCAGGGTGTCTATTAAGTTTGTCTATATGTGGCATTTGGGCAATTTTTC 1018
QY 844 TAGTTCCAATTAATTCGTAGAGATGTGGACCGGCTATTTTGTCTCAGAAAACGATTTGGAC 903
Db 1019 TAGTTCCACAAATCAGAAAGATGTGGACCGGCTATCTTTCTCAAATAGTAGGTCT 1078
QY 904 AGAATGGACGATATTTTAATCTACAAGTTTCGATCGATGTATGTGTGATGCTGAGGAGC 963
Db 1079 GTAATGGTAGGATTTTGTAGATCTATAAATTCAGATCAATGCGAGTAGATCAGAACAA 1138
QY 964 GCAAAAAGACCTGCTCAGCCAAACAGATCAGAGGTGGGTATGTTTAAATGGGAA 1023
Db 1139 TTAAGAAAGATTAATAGTTTCAAAATCAAAATGACAGG-----CTAATGTTTAAAGTTAGA 1193
QY 1024 AAACGATCTAGAAATPACTCCAATTCGACATTTTATACGCAAAAAACAAGTTTAGACGAGT 1083
Db 1194 CGATGATCTAGAAATTAATAATAGGAAAAATTTATTTCG-AAAACAAGCATAGATGAGT 1252
QY 1084 TACCACAGTTTTTAATGTTTTTAATTTGGCGATATAGTCTAGTTGGTACACGTCACCTTA 1143
Db 1253 TGCCTCAATCTATATATGTTTTAAAAAGGTGATATGAGTTTGTAGGAAACACGCCCTCCCA 1312
QY 1144 CAGTTGATCAATTTGAAAATATACCTCGGTCAAAAGAGACATGTAGTTTTAAACCG 1203
Db 1313 CAGTTGATGAATATGAAAAGTAAATTCACGCAAGAGCGCGCTTAGTTTAAAGCCAG 1372
QY 1204 GGATTAACAGTCTCTGGCAGGTGTAGTGTGCTAGTAAATATACAGACTTCGACGAGGTAG 1263
Db 1373 GAATCACTGTTTGGGCAAAATATCTGGTAGAAATATATCACTGATTTTGATGAATCG 1432
QY 1264 TTCGGTTGCACTTAGCATATGATGATTAATTTGGAATATCTGGTCAGATATTAATAATTTTAT 1323
Db 1433 TAAAGTTAGATGTTCAATATATCAATGAATGCTCTATTTGGTCAGATATTAAGATTAATTC 1492
QY 1324 TAAAGCAGTGAAGTTGTTATTTGTGAGAGGGAGTAAAGTAAAGTA----- 1372
Db 1493 TCCTAACACTAAGGTAGTCTTACTTGGACAGGTGCTTAAGTAAAGGTAAAGTTTGAAG 1552
QY 1373 -----TATGAAGTTTGTGTCGGTTCTTCAGGGGACATTTGACTCACTTGTATTT 1426
Db 1553 GAATATAATGAAAAATTTGCTGTTGTTCAAGTGTGTCTATCTAGCACACTTGAACCT 1612

QY 1427 GTTAAAAACGTTTTTGGAAAGGAAGAACGTTTTTTGGGTAAACATTTTGATAAAGAGATGC 1486
Db 1613 TTTGAAACCCATTTTGGGAAAAAGAGATAGGTTTTTTGGGTAAACCTTTTGATAAAGAGATGC 1672
QY 1487 AAGAAAGTCTTTTGAAGAATGAAAAAATGTATCCATGTTTACTTTTCCAAACAAATCGCAATCT 1546
Db 1673 TAGGAGTATCTAAGAGAGAGATTTGTATATCATTTGCTTCTTTCCAAACAAACCGTAAATGT 1732
QY 1547 CATTAATTTAGTGAATAATACCTTTCTTAGCTTTTCAAAATTTTTCGTTGATGAGAAACCGA 1606
Db 1733 CAAAAACCTTGTGTAATAATACTATTTCTAGCTTTTAAAGTCTTTAGAAAAAGAAAGACCGA 1792
QY 1607 TGTATATTTTCACTCTGGTGGCGCTGCTGCTGCCCTTTCTTTTATCATCGGAAAAACTATT 1666
Db 1793 TGTATATCATATCATCTGGTGGCGCTGTAGCAGTACCAATCTTTTATATTTGTTAAAGTTATT 1852
QY 1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTATTAATCTACATTAACCTCG 1726
Db 1853 TGGTTGTAAGACCGTTTATATAGAGTTTTCACAGGATAGATAAACCAACTTTTGACAGG 1912
QY 1727 AAAACTAGTTTATCCCGTAAACAGATATTTTATTTCTCAGTCGGGAAGAAATGAAAGAGT 1786
Db 1913 AAAATTAGTGTATCTGTAAACAGATAAATTTATTTCTCAGTCGGGAAGAAATGAAAGAGT 1972
QY 1787 ATATCCTAAATCTATAAATTTGGGGAGTATTTTTTAATGATTTTTGTACACAGTAGAACT 1846
Db 1973 TTATCCTAAGGCAATTAATTTTAGGAGGAATTTTTTAAATGATTTTTTGTACAGTGGGACA 2032
QY 1847 CATGAACAACAGTTTAAATCGATTTGATTAAGAGATTTGATTTATGAAAAAATGGAAGT 1906
Db 2033 CATGAACAGCAGTTTCAACCGTCTTATTAAGAAAGTTGATTAAGAAAGGACAGTGTCT 2092
QY 1907 ATAAACCGAGAAATTAATTTCAAAACAGGATATTTCTGACTATATATTTCCAGAAATTTGCAAG 1966
Db 2093 ATTGATCAAGAAGTGTTCATTCAAAACGGTTACTCAGACTTTGAACTTCAGAAATTTGTGAG 2152
QY 1967 TATAAAAAATTTCTCAGTTTCAAAAGAAATGGAAACAATATATTAACAATCAGAAAGTAGTT 2026
Db 2153 TGGTCAAAATTTCTCTCATATGATGATGAACTCTTACATGAAAGAGCTGAGATTTGT 2212
QY 2027 ATTTGCCAGGAGGCGCCGCTACTTTTATGATTTTATGATTTATCCAAAGGAAAAACAATTA 2086
Db 2213 ATCACATATGGCGGTCCAGCGACGTTTATGAATGAGTTTCTTAAAGGGAAAAACAATTA 2272
QY 2087 TTGTTTCTTAGACAAAAAAGTATCGTGAAACATGTAAATGATCATCAAGTAGAGTTTGT 2146
Db 2273 GTGGTTCTTAGACAAGACAGTTTGGAGAGCATGTGAATATCATCAGGTGATTTTGT 2332
QY 2147 AGAAGAAATTTTACAAGATAATAATATTTTATTTATAGAAAAATATAGATGATTTGTTGAA 2206
Db 2333 AAAGAGTTTATTTCTTGAATATGAATTTAGATTTATATTTTGAATATCATGTAATTTAGAGAT 2392
QY 2207 AAAATTTTGAAGTTTCTAAGCAAC---TAACTTTTATCAATAATAATTAATTTTGT 2263
Db 2393 ATTATTAGGAAAAAATAATATCTACTAGTAAGTAATATCAAAAAACAATGATTTTGT 2452
QY 2264 GAAAGATTTAAACAAATAGTTTGAATAATTTAATGAGGATCAAGAAAAATGAAATAATAAAAA 2323
Db 2453 TCCCTTTCAAAATGAACCTTTCAACTATTTGATTAATAATATATTTTGTGAGAAAAA 2512
QY 2324 AGATGCA 2330
Db 2513 AATTGAA 2519

RESULT 18
AF363057
LOCUS
DEFINITION
Streptococcus agalactiae strain GB00/017 capsular polysaccharide
bioynthesis gene cluster, partial sequence.
ACCESSION
AF363057
VERSION
AF363057.1 GI:18029155

| | | | | |
|----------------------------|-----------------------------------------------------------------|-----------------------------------------------------------|-------------|----------------------------------------------------------|
| KEYWORDS | Streptococcus agalactiae | gene | 2587..3726 | ONCWSKPLSYDDMNSYNKEAEIVITHGGPATFMNAVSKGKKTIVVPQEQFGHHNN |
| SOURCE | Streptococcus agalactiae | CDS | 2587..3726 | HQVDFLKLFLKVELDYILNISELENIKERNIISTKVISQNDPFCSSFKNELSKLFE |
| ORGANISM | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; | | | |
| REFERENCE | Streptococcus | | | |
| AUTHORS | 1 (bases 1 to 4888) | | | |
| TITLE | Kong,F., Gowan,S., Martin,D., James,G. and Gilbert,G.L. | | | |
| | Serotype identification of group B streptococci by PCR and | | | |
| | sequencing | | | |
| JOURNAL | J. Clin. Microbiol. 40 (1), 216-226 (2002) | | | |
| MEDLINE | 21635022 | | | |
| PUBMED | 11773119 | | | |
| REFERENCE | 2 (bases 1 to 4888) | | | |
| AUTHORS | Kong,F. and Gilbert,G.L. | | | |
| TITLE | Direct Submission | | | |
| JOURNAL | Submitted (22-Mar-2001) CIDMLA, ICPMR, Westmead Hospital, Darcy | | | |
| | Road, Sydney, NSW 2145, Australia | | | |
| FEATURES | Location/Qualifiers | | | |
| source | 1..4888 | gene | 3726..4700 | TLTITLFSFALLELDGANWLVIFTVLGILENDKDFYSQLKRWKS" |
| | /organism="Streptococcus agalactiae" | | | |
| | /mol_type="genomic DNA" | | | |
| | /strain="GB00/017" | | | |
| | /serotype="1a" | | | |
| | /db_xref="taxon:1311" | | | |
| gene | <1..135 | | | |
| CDS | /gene="cpsD" | | | |
| | <1..135 | | | |
| | /gene="cpsD" | | | |
| | /codon_start=1 | | | |
| | /transl_table=11 | | | |
| | /product="Cpsd" | | | |
| | /protein_id="AAL56374.1" | | | |
| | /db_xref="GI:18029156" | | | |
| | /translation="AKEQMEQSGSKFLGIILNKVSESVAITYGDIYGYNGYKRDREK" | | | |
| gene | 361..1536 | gene | 4734..>4888 | REIIFDKIKSPFYKAVKYLKGLLSFYLKMCSPKLYVMAYRFPQK" |
| CDS | /gene="cpsE" | | | |
| | 361..1536 | | | |
| | /gene="cpsE" | | | |
| | /codon_start=1 | | | |
| | /transl_table=11 | | | |
| | /product="CpsE" | | | |
| | /protein_id="AAL56375.1" | | | |
| | /db_xref="GI:18029157" | | | |
| | /translation="WLVKVSFYIPISSLPFFIPKNSPTTTLRSLGFFSFIAMNLSLLYL | | | |
| | LSNFKYKYSYAKFSRDTKVLTITNKSLSKMTFRNKYDHNVIACVILDSSEKCY | | | |
| | DKNSLRIRNKALTSCLTVDQAFINIPFLFKYIQIDINDIENGVIYVNV | | | |
| | VEALSPNIGSKRIQTFEGYSVITYSMKFYKYSHLIAKRFIDITGAILIGLCIGVAI | | | |
| | FLVPQIRKDGPAIFSQNRVGRNRIFRFYPSNRVDAEQIKDILLVHNMQTMGLMK | | | |
| | LDDPRITKIGKFIKRTSIDELPOFYNVLKGDMSLVGRPTTVDYEKYNSGTQRRLS | | | |
| | FKPGITGLWQISGRNITDFDEIVKLDVQYINWSIWSIDIKIILLTLKVLVLTGAK" | | | |
| gene | 1560..2009 | | | |
| CDS | /gene="cpsF" | | | |
| | 1560..2009 | | | |
| | /gene="cpsF" | | | |
| | /codon_start=1 | | | |
| | /transl_table=11 | | | |
| | /product="CpsF" | | | |
| | /protein_id="AAL56376.1" | | | |
| | /db_xref="GI:18029158" | | | |
| | /translation="MKICLVSGSGGHLAHLNLLPIKEWEDRFWVTFDKEDARSILRE | | | |
| | BIVYHCFPTNRNKNLVKNTILAPKVLKERPDVLISSGAVAVPPFYVIGLFGCKT | | | |
| | VYIEVFDRIQTLTGKLVYPVTDKFIYQWHEMKVYPKAINLGIF" | | | |
| gene | 2009..2491 | | | |
| CDS | /gene="cpsG" | | | |
| | 2009..2491 | | | |
| | /gene="cpsG" | | | |
| | /codon_start=1 | | | |
| | /transl_table=11 | | | |
| | /product="CpsG" | | | |
| | /protein_id="AAL56377.1" | | | |
| | /db_xref="GI:18029159" | | | |
| | /translation="WIFVTVGTHEQQFNRLIKEVDRLKGTGAIQDEVFIQTGYSDPEP | | | |
| gene | 3161..4888 | gene | 4889..5136 | QKFNESLDFTTFIENEVSSIVPAKLANYVEAKFLREKIKLRKMFELGSDNKKIKVQ |
| CDS | /gene="cpsH" | | | |
| | 3161..4888 | | | |
| | /gene="cpsH" | | | |
| | /codon_start=1 | | | |
| | /transl_table=11 | | | |
| | /product="CpsH" | | | |
| | /protein_id="AAL56380.1" | | | |
| | /db_xref="GI:18029162" | | | |
| | /translation="MIKKIEKDLISVIPIYVNVEDYLVECIIESLIVQTYRNIIEILLN | | | |
| | DGSTDNCA" | | | |
| ORIGIN | | | | |
| Query Match | 10.4%; Score 727.8; DB 1; Length 4888; | | | |
| Best Local Similarity | 59.2%; Pred. NO. 3.9e-73; | | | |
| Matches 1389; Conservative | 0; Mismatches 917; Indels 41; Gaps 7; | | | |
| Qy | 16 | TGGCAATATTGATGATAGCAGTTGCAATTTCTGCAATCTTACAAAGTCATATACCAA | 75 | |
| Db | 182 | TAGCGATGATTCAACACAGTTGTGGTTTTATTTCTGCAAGTTTGCAATTAACATTA | 241 | |
| Qy | 76 | ATGCTGATTTAAATCG---TTCTGCAATTTTATCATAATGATGTTTCATTTATTTG | 132 | |
| Db | 242 | CTCCCAATTTAAAGCAATAAGATTATTGTTGTTCTATTGATACATATATTGCT | 301 | |
| Qy | 133 | TTTTTATATCTCGTATGCCAGTTGAAATTTAGTATAGAGGTAATCTGATAGATTG | 192 | |
| Db | 302 | TTTATCTTCTGATTTTACAGACACTTTTGGAGTCGTGGCTATCTTGAAGAGTTT | 361 | |
| Qy | 193 | AAACATTTAACTAGTATAATAATTGCAATTTTCTTACGGCAGTATCATTTTGTGG | 252 | |
| Db | 362 | TGGTATTGAAATACAGCTTTTACTATATTTTACATCAAGTTTCATTTATTTT | 421 | |
| Qy | 253 | AGATAATTTTCGCACTTTCAAGACGTGTCGCGGTATTTCACATTAATAAATTCG | 312 | |
| Db | 422 | AAAATCATTTTACACGACGACTTTCCTTTTTTTCTTTTATTCCTATGAATTCG | 481 | |
| Qy | 313 | TGGTATACCTATTAAACCGTAATTATTAAAGCAGTTTAAAGATAGCTTCTATT | 372 | |

Db 482 TATTGTATCTATTGAATTCATTTTAAATAATATCGAARATATCTTACGCTAAGTTTT 541
Qy 373 TCTATCAAAAAGACAGATCTTAATACACGGCTGAAACGATGGGAAAAATATGCAAGTTT 432
Db 542 CACGAGATACCAAAAGTTGTTTGTATAACGATAGGATCTTTTATCAAAAATGACCTTTA 601
Qy 433 TATTGTAATCACATAAACAATTCAAAATAATCTGTTGTCATGCTAGTTTAAAGTACAG 492
Db 602 GGAATAAAT---ACGACCAATAATATATGCTGCTGATCTTGGACTCTCTGAAAGG 658
Qy 493 AAATAGATAAAATTAATTTATCATTTACCGCTCTATTATCTGTGGAAGAAGCTATAGAGT 552
Db 659 ATTGTTATGATTGAAACATAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 718
Qy 553 TTTCAACAGGGAAGTGGTCGACACAGCTCTTTTAAATCTACCAAGTGAGTTTTTAGACG 612
Db 719 AGTTAAGCTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCAATGGAATTTTGGTA 778
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTGAAGTTGTTAGGTATTGATGTAAGCGTTG 663
Db 779 AATACCAATAACAAGATAATTAATGACATTGAAGCAATGGGAGTGATTTGTCATGTTA 838
Qy 664 ATATTAAATTCATTCCGTTTACTCGTTTGAACCAAAAAATCCAACTGCTAGGTGACC 723
Db 839 ATGTAGAGGCACTTAGCTTTGATTAATATAGGAGAAAGCGAATCCAACTTTTGAAGAT 898
Qy 724 ATAGCATGTAACTTTTCCAAATTTTATAAGCCTAGTCATATCATGATGAAACGAC 783
Db 899 ATAGTGTATTACATAATCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT 958
Qy 784 TTTTGGATATACCTCGAGCGGTAGTCGGTTAAATTTATTTGGGTATAGTTTCTATTGTT 843
Db 959 TTTTGGATATACCGGGTGTATTATAGTTTGTCTATGTGCAATGTGGCAATTTTTC 1018
Qy 844 TAGTTCCAAATTAATTCGTAGATGGTGGACCGGCTATTTTGTCTCAGAAACGAGTTCGAC 903
Db 1019 TAGTTCCAAATCAGAAAGATGGTGGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC 1078
Qy 904 AGAATGGACGATATTTACATCTTACAAGTTTTCGATCGATGATGTTGATGCTCAGGAGC 963
Db 1079 GTAATGGTAGGATTTTATAGATTCTATAAATTCAGATCAATCGAGTAGATCGACAAACA 1138
Qy 964 GCRAAAGACTCTCGACCAAAACAGATGCAAGGTGGGTATGTTTAAATGGAA 1023
Db 1139 TTAAGAAAGATTTATAGTTCACAATCAAAATGACAGGG-----CTAATGTTTAAAGTTAGA 1193
Qy 1024 AAACGATCCTAGAATTAATCTCAATTTGGACATTTTCATACGCCAAAAACAAGTTTACAGAGT 1083
Db 1194 CGATGATCCTAGAAATTAATAAATAGGAAATTTATTTCG-AAAAACAAGCATAGATGAGT 1252
Qy 1084 TACCACAGTTTATTAATGTTTAAATTTGGCGATATGAGTCTAGTTGGTACAGCTCCACTA 1143
Db 1253 TGCCTCAATTTCTATAATGTTTTAAAGGTGATAGAGTTTATAGTAGGACACAGCCCTCCCA 1312
Qy 1144 CAGTTGATGATTTGAAAAATATCTCTGCTGCTCAAAAGAGACGATTCGAGTTTAAACAG 1203
Db 1313 CAGTTGATGATATGAAGAATTAATTTCAACGACAGAGCGACGCTTAGTTTAAAGCCAG 1372
Qy 1204 GGATTACAGGTCTCTGGCAGGTAGTGGTCTGCTAGTAAATATACACAGCTTCGACGACGTAG 1263
Db 1373 GAATCAGCTGGTTTGTGGCAATATCTGGTAGAATATATACATGATTTTGTGATGAATCG 1432
Qy 1264 TTCGGTTGGACTTAGCATACATGATTAATTTGGACTATCTGGTCAGATTAATAAATTTTAT 1323
Db 1433 TAAAGTTAGATGTTTCAATATATCAATGAATGCTCTATTGCTCAGATTAAGATTTATTC 1492
Qy 1324 TAAAGACAGTCAGAGTTGTTTGTGAGAGGAGGAGTAAAGTA----- 1372
Db 1493 TCCTAAACCACTAAGAGTAGTCTTACTTTGGGA CAGGTGCTTAAGTAAGGTTGAAG 1552
Qy 1373 -----TATGAAAGTTTCTTTGGTCCGTTCTTTCAGGGGGACATTTGACTCACTTGTATTT 1426
Db 1553 GAATATAATGAAATTTGCTCTGGTTGGTTTCAAGTGGTGGTCTATCTAGCACACTTGAACCT 1612

Qy 1427 GTTAAACCGTTTTTGGAGGAAGAACGTTTTTGGGTAAACATTTTCATAAAGAGGATGC 1486
Db 1613 TTTGAAACCCATTTTGGGAAAAAGAGATAGTTTTTGGGTAACTTTTGTATAAAGAGATGC 1672
Qy 1487 AAGAAGTCTTTTCAAGAAATGAAAAAATGTATCATGTTACTTTTCCAAACAATCCCAATCT 1546
Db 1673 TAGAGATTTCTAAGGAGAGATTTGTATATCATTTGCTTCTTTCCAAACAACCCGTATGT 1732
Qy 1547 CATTAATTTTAGTGAAAAATACTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606
Db 1733 CAAAAACTTGGTAAAAAATACTATTTCTAGCTTTTAAAGTCTCTTAGAAAAAGAAAGACGAGA 1792
Qy 1607 TGTATTATTTTCACTGCTGGCGCGTTCGTCTCCCTTCTTTTACATCGGAAACTATT 1666
Db 1793 TGTATTATCATATCTGCTGGTCCGCTGTAGCAGTACCAATCTTTTATATTTGGTAAAGTTATT 1852
Qy 1667 TGGAGCAAGACGATTTATATTCGAAGTATTATTCGAGTTTAAATAAATCTACATTAACCTGG 1726
Db 1853 TGGTTGTAGACCGTTTATATAGAGGTTTTCGACAGATAGATATAACCAACTTTGACAGG 1912
Qy 1727 AAACTAGTTTATCCCGTAAACAGATATTTTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGGT 1786
Db 1913 AAAATTAGTGTATCTCTGTAACAGATAAATTTATTTGTTTCAGTGGGAAGAAATGAAAAAGT 1972
Qy 1787 ATATCTTAATCTATTAACTTTGGGAGTATTTTTTAATGATTTTGTAAACAGTAGGAACT 1846
Db 1973 TTATCTTAAGGCAATTAATTTAGGAGGAATTTTTTAATGATTTTTTGTTCACAGTGGGACA 2032
Qy 1847 CATGAAACAACAGTTTAAATCGATTGATAAAAGAGATTGATTATTTGAAAAAATAATGGAAGT 1906
Db 2033 CATGAAACGCGTTTCAACCGCTTTTATTAAGAGAGTTGATAGATTAAAGGACAGGTGCT 2092
Qy 1907 ATAACCGACGAAATATTTTATCAACAGAGTATTTCTGACTATATTTCCAGAAATATTGCAAG 1966
Db 2093 ATTGATCAAGAAGTGTTCATTCAACCGGTTACTTCAGACTTTTGAACCTTCAGAAATTGTCAG 2152
Qy 1967 TATAAAAAATTTCTCAGTTTACAAAGAAATCGAAACAATATATTATTAACAATCAGAAGTAGTT 2026
Db 2153 TGGTCAAAATTTCTCTCATATGATGATGAACTCTTTACATGAAAGAGCTGAGATTGTT 2212
Qy 2027 ATTTGCCACGAGGCCCGCTACTTTTATGAAATTCATTTATCCAAAGGAAAAACAATTA 2086
Db 2213 ATCACAATGCGCGTCAGCGAGCTTTATGAAATGCGAGTTTCTAAAGGGAAAAACAATATT 2272
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGTGA 2146
Db 2273 GTGGTTCCTAGACAAAGAACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGAATTTTGTG 2332
Qy 2147 AGAAGATTTTACAGATAATAATATTTTATTTATAGAAAAATATAGATGATTGTTGTGA 2206
Db 2333 AAGGAGTTATTTCTTGAATAATGAATTAGATTATATTTTGAATATTCAGTGAATTAGAGAA 2392
Qy 2207 AAAATTATTCGAAGTTTCTAAGCAAAAC---TAACTTTTACATCAAAATAATAATTTTTTTGT 2263
Db 2393 ATTATTAGGAAAAAATAATATCTACTAGTAAGATATATACAAAAACAATGATTTTTTGT 2452
Qy 2264 GAAAGATTAAAAACAATAGTTGAAAAATTTTAAATGAGGATCAAGAAATATGAATAATAAAA 2323
Db 2453 TCCTCTTTCAAAAATGAACTTTCTAAACTATTTTGAATAAATAATATTTTGTGGAGAAAAA 2512
Qy 2324 AGATGCA 2330
Db 2513 AATTGAA 2519

RESULT 19
AF363059
LOCUS
DEFINITION
Streptococcus agalactiae strain SG99/276 capsular polysaccharide
biosynthesis gene cluster, partial sequence.
AF363059
AF363059
VERSION
AF363059.1 GI:18029171

| | | | |
|----|------|-----------------------------------------------------------------|------|
| Db | 482 | TATTGTATCTATTGAAATTCATTTTAAAAATATTATCGAAAAATATCTTACGCTAAGTTTT | 541 |
| Qy | 373 | TCTATCAAAAAAGACGATTTCTAATTACAACGGCTGAACGATGGAAAAATATCGACAGTTT | 432 |
| Db | 542 | CACGAGATACCAAGTTGTTTGTATAACGATAAGGATTTCTTTATCAAAATGACCTTTA | 601 |
| Qy | 433 | TATTGGAATCACATAAACAATTCAAAAAATCTTTGTTGCAATGTGTAGTTTTAGGTACAG | 492 |
| Db | 602 | GGAAATAAT---ACGACCAATAATTATATGCTGTCTGCATCTTGGACTCTCTCGAAAAAG | 658 |
| Qy | 493 | AAATAGATAAAATTTAAATTTATCATTCGGCTCTATTATTTCTGTGGGAAGACTATAGAGT | 552 |
| Db | 659 | ATTGTTATGATTTGAAACATAAATCGTTAAGGAATAAFAAACAAAGATGCTTCTTACTTCAG | 718 |
| Qy | 553 | TTTTCAACAGGGAGTGGTCGACACAGCTTTTATAAATCTACCAAGTGAGTTTTTAGAAG | 612 |
| Db | 719 | AGTTAACTGCTTAACTGTTGATCGAGCTTTTATTAAACATACCCATTTGAATTAATTTGGTA | 778 |
| Qy | 613 | TAAAGCAATTC-----GTTTCAGATTTTGAGTTTGTAGGTATTGATGTGAAGCGTTG | 663 |
| Db | 779 | AATACCAAAATACAAGATATTATTAAATGACATGAAGCAATGGGAGTGATTTGTCAATGTTA | 838 |
| Qy | 664 | ATATTAAATCAATTCGGTTTTACTCGGTTGAAAAACAAAAAATCCAACTGCTAGGTGAC | 723 |
| Db | 839 | ATGTAGAGGCATTTAGCTTTTGTATAATATAGGAGAAAAACGAATCCAACTTTTGAAGGAT | 898 |
| Qy | 724 | ATAGCATTTGAACTTTTTTCCACAAATTTTTATTAAGCGCTAGTGCATATCATGATGAACGAC | 783 |
| Db | 899 | ATAGTGTTATTACATAATCTATGAAATTCATATAATATAGTACCTTATAGCAAAACGAT | 958 |
| Qy | 784 | TTTTGGATATACTCGGACGGTAGTCGGGTTAAATTTATTTCTGGTATAGTTTCTATTTTGT | 843 |
| Db | 959 | TTTTGGATATCACGGTTGCTATTATAGTTTGCTCATATGTGGCAATGTGGCAATTTTTC | 1018 |
| Qy | 844 | TAGTTCCAAATTAATTCGTAGAGATGGTGGACCGGCTATTTTTGCTCGAAAAACGAGTTGGAC | 903 |
| Db | 1019 | TAGTTCCACAAATCAGAAAAAGTGGTGGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC | 1078 |
| Qy | 904 | AGATGACGCAATATTTACATCTTACAGTTTCGATCGATGTATGTTGATGCTGAGGAGC | 963 |
| Db | 1079 | GTAATGTGTAGGATTTTGTAGATTTCTATAAATTCAGATCAATTCGCGAGTAGATGCGAACA | 1138 |
| Qy | 964 | GCAAAAAAGACTTTGCTCAGCCMAAACAGATGCAAGGGTGGGTATGTTTTAAAAATGGAA | 1023 |
| Db | 1139 | TTAAGAAAGATTTATTAGTTTCAATCAATGACAGGG-----CTAATGTTTAAATTAGA | 1193 |
| Qy | 1024 | AAACGATCTAGAAATTAATCTCAAATGGACATTTTCATACGAAAAACAAAGTTTACAGAGT | 1083 |
| Db | 1194 | CGATGATCTTAGAATTAATAAAATAGGAAAAATTTATTCG-AAAAACAAGCATAGATGAGT | 1252 |
| Qy | 1084 | TACCACAGTTTTATAATGTTTTAATTTGGCGATATAGTCTAGTTGGTACAGTCCACCTA | 1143 |
| Db | 1253 | TGCCTCAATTTCTATAATGTTTTAAAGAGTGATAGTGTAGTAGGAAACAGCGCCTCCCA | 1312 |
| Qy | 1144 | CAGTTGATGAATTTGAAAAATATATCTCCTGGTCAAAAGAGACGATTCAGTTTTTAAACCG | 1203 |
| Db | 1313 | CAGTTGATGATATGAAGAATTAATTTCAACGACGAGACGAGCCTTAGTTTTTAAGCCAG | 1372 |
| Qy | 1204 | GGATTAACAGTCTCTCGCAGGTTAGTGGTGGTAGTAATATCACAGACTTCGACGACGTAG | 1263 |
| Db | 1373 | GAATCACTGGTTTTGTGGCAAAATATCTGGTAGAAAAATATATCACTGATTTTGTAGTAATCG | 1432 |
| Qy | 1264 | TTCCGTTGGACTTAGCATACATTCGATAATTTGGACTATCTGGTCAGATATTAAAAATTTTAT | 1323 |
| Db | 1433 | TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTGGTCAGATATTAAAGATTATTC | 1492 |
| Qy | 1324 | TAAAGACAGTCAAAAGTTGTTATTTGTAGAGAGGGAAGTAAAGTAAAAAGTA----- | 1372 |
| Db | 1493 | TCCTAAACAATAAGGTAGTCTTACTTCGGGACAGTGTCTAAGTAAAGGTAGGTTTGAAG | 1552 |
| Qy | 1373 | -----TATGAAAGTTTTGTTGGTCGGTTCTTTCAGGGGGACATTTGACTCACTGTGATTT | 1426 |
| Db | 1553 | GAATATAATGAAAAATTTGCTCTGGTTGGTTTCAAGTGGTGGTCACTACGACACTTGAACCT | 1612 |

| | | | |
|-----------------|-------------------------------------------------------------|--------------------------------------------------------------------|--------|
| Qy | 1427 | GTTTAAACCCGTTTTTGAAGGAAGAAAGCGTTTTTTGGGTAAACATTTTGTATAAAGAGGATGC | 1486 |
| Db | 1613 | TTTTGAAACCCATTTTGGGAAAAAGAGATAGGTTTTTGGGTAAACCTTTTGATAAAGAGATGC | 1672 |
| Qy | 1487 | AAGAAGTCTTTTGAAGAATGAAMAAATGTATCCATGTTACTTTTCCAAACAAATCGCAATCT | 1546 |
| Db | 1673 | TAGGAGTATTTCTAAGAGAGAAGATGTCATATCATTTGCTTCTTTTCCAAACAAACCGTAAATGT | 1732 |
| Qy | 1547 | CATTAAATTAGTCGAAAAATACTTTTCTTAGCTTTTCAAAATTTTACGTCGATGAGAAACCAGA | 1606 |
| Db | 1733 | CAAAAACTTGGTAAAAAATACTATCTCTAGCTTTTAAAGGTCCTTTAGAAAAAGAAAGACAGA | 1792 |
| Qy | 1607 | TGTTATTTATTTCAATCTGGTTCGGCCGCTGCTGTCCCTTCTTTTACATCGGAAAACTATT | 1666 |
| Db | 1793 | TGTTATCATATCATCTGGTTCGGCCGCTAGCAGGTACCATTTCTTTTATTTAGTAACTTATT | 1852 |
| Qy | 1667 | TGGAGCAAGACGATTTATATTTGAAGTATTTTGATCGAGTTAATAAATCTACATTAACCTGG | 1726 |
| Db | 1853 | TGGTTGTGAACCGTTTTATATAGAGGTTTTTCGACAGGATAGATAAAACCACTTTGACAGG | 1912 |
| Qy | 1727 | AAAACCTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAGTCGGGAGGAATGAAGAAGCT | 1786 |
| Db | 1913 | AAAATTAGTGATCTCTGTAAACAGATAAATTTATTTGTTTCAGTCGGGAGGAATGAAAAAGT | 1972 |
| Qy | 1787 | ATATCTTAAATCTATTAACTTTGGGGAGTATTTTTTAAATGATTTTTTGTAAACAGTAGGAACT | 1846 |
| Db | 1973 | TTATCTTAGGCAATTAATTTTAGAGGAATTTTTTAATGATTTTTTGTAAACAGTAGGAACT | 2032 |
| Qy | 1847 | CATGAACAAACAGTTTTAATCGATTTGATTAAGAGAGATGATTTTATTTGAAAAAATAATGGAAGT | 1906 |
| Db | 2033 | CATGAACAGCAGTTCAACCGTCTTATTAAGAAGTTGTATAGATTTAAAGGGACAGGTGCT | 2092 |
| Qy | 1907 | ATAACCGAGAAATATTTATTTCAAACAGAGATATTTCTGACTATATTTCCAGAAATTTTCAAG | 1966 |
| Db | 2093 | ATTGATCAAGAAGTGTTTCATTTCAAACGGGTTACTCAGACTTTTGAACCTCAGAAATTGTCAG | 2152 |
| Qy | 1967 | TATAAAAAATTTCTCAGTTTACAAGAAAAATGGAAACATATATTTAAACAAATCAGAAAGTAGTT | 2026 |
| Db | 2153 | TGGTCAAAATTTCTCTCATATGATGATGAACTCTTTCATGAAAGAGCTGAGATTTGTT | 2212 |
| Qy | 2027 | ATTTGGCCACGGAGGCCCGCTACTTTTTATGAATTCATTATCCAAAGGAAAAAACAATTA | 2086 |
| Db | 2213 | ATCACACATGGCGGTCCAGCGAGCTTTATGAATGCAGTTTCTAAAGGGAAAAAACTATT | 2272 |
| Qy | 2087 | TTGTTTCTAGACAAAAAAGTATGTTGGAACATGTAAATGATCATCAAGTAGAGTTTGTA | 2146 |
| Db | 2273 | GTGGTTCTTAGACAAGAACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGATTTTTTG | 2332 |
| Qy | 2147 | AGAAGAATTTTACAAGATAATAATTTTTATTTATAGAAAAATATAGATGATTTGTTTGA | 2206 |
| Db | 2333 | AAGAGTTATCTTGAATAATGAATTAGATTAATTTTGAATATCAGTGAATTAGAGAT | 2392 |
| Qy | 2207 | AAAATTATGAAGTTTCTTAAGCAAC---TAACTTTACATCAAAATAATATTTTTTTTGT | 2263 |
| Db | 2393 | ATTATTAAAGCAAAAAATATATCTACTAGTAAAGTAAATATCACAAAACAATGATTTTTGT | 2452 |
| Qy | 2264 | GAAAGATTAAAACAAATAGTTTGAAAAAATTTAATGAGGATCAAGAAAAATGAATTAATAAAA | 2323 |
| Db | 2453 | TCCTCTTTCAAAAATGAACCTTTCTAAACTATTTTGAATAAATAATATTTTTTGTGGAGAAAA | 2512 |
| Qy | 2324 | AGATGCA | 2330 |
| Db | 2513 | AAATTGAA | 2519 |
| RESULT 20 | | | |
| BD166249 | | | |
| LOCUS | BD166249 | 6865 bp | linear |
| DEFINITION | Beta-1,3-galactosyltransferase and DNA encoding the enzyme. | | |
| ACCESSION | BD166249 | | |
| VERSION | BD166249.1 | | |
| KEYWORDS | JP 2002199885-A/2. | | |
| PAT 17-JAN-2003 | | | |

SOURCE unidentified
ORGANISM unidentified
REFERENCE unclassified.
1 (bases 1 to 6865)
AUTHORS Miyake, K., Watanabe, M. and Iijima, S.
TITLE Beta-1,3-galactosyltransferase and DNA encoding the enzyme
JOURNAL Patent: JP 2002199885-A 2 16-JUL-2002;
KYOWA HAKKO KOGYO CO LTD
OS Streptococcus agalactiae Type Ib
PN JP 2002199885-A/2
PD 16-JUL-2002
PF 05-JAN-2001 JP 2001000392
PI KATSUHIRO MIYAKE, MASAKI WATANABE, SHINJI IIJIMA PC
C12N15/09, C12N15/21, C12N9/10, C12P19/28//C12N15/09, C12R1:46), PC
(C12N9/10, C12R1:19), (C12P19/28, C12R1:19), (C12N15/00, (C12N15/00, PC
C12R1:46)
CC Beta-1,3-galactosyltransferase and DNA encoding the enzyme FH
Key Location/Qualifiers
FT CDS (617).. (1789)
FT CDS (1816).. (2262)
FT CDS (2265).. (2744)
FT CDS (2843).. (3979)
FT CDS (3982).. (4953)
FT CDS (5009).. (5947).
FEATURES
source
1..6865
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
ORIGIN
Query Match 10.4%; Score 727.8; DB 6; Length 6865;
Best Local Similarity 59.2%; Pred. No. 3.5e-73;
Matches 1389; Conservative 0; Mismatches 917; Indels 41; Gaps 7;
Qy 16 TGGCAATTAATGATAGATAGCAGTTGCAATTTCTGCAATCTTAAGATCATATACCAA 75
Db 438 TAGCGATGATTCAAACCGTTGGTATTATTTCTGCAAGTTTGACATTAACATTAATTA 497
Qy 76 ATGCTGATTAATATCG---TTCTGGAATTTTATCATATGATGTTTCATTTATTTGCA 132
Db 498 CTCCTCAATTTTAAAGCAATTAAGATTTATTGTTGTTGTTCTATGATATATATTTGCT 557
Qy 133 TTTTATATATCTGATGCGAGTTGAAATTTGAGTATAGAGTAATCTGATAGAGTTTGAAA 192
Db 558 TTTATCTTCTGATTTTACAGAGACTTTTGAGTCTGGCTATCTTGAAGTTTAA 617
Qy 193 AAACATTTAATAGTATAATTTGCAATTTTCTTAAGCAGTATCATTTTGTGG 252
Db 618 TGGTATTGAATATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTCTA 677
Qy 253 AGAATAATTTGCACTTTCAAGACGTTGGTGGCTGATTTTCATATTAATTAATCTGTT 312
Db 678 AAAACTCTTTTACACGACAGCACTTTCTTTTACTTTTATGCTATGAATTCGATTT 737
Qy 313 TGGTATACCTATTAAACGTAATTTAAGCAGTTTAAAGATAGCTTTCTATTTCGACAA 372
Db 738 TATTATATCTATGAAATTCATTTTAAATATATATCGAAATATTTCTTACGCTAAGTTT 797
Qy 373 TCTATCAAAAAGAACGATTTCTAATTTACAAACGCTTGAACGATGGGAAATATGCAAGTTT 432
Db 798 CACGAGATACCAAGCTGTTTGTGATAACGAATAAGGATTTCTTATCAAAAATGACCTTTA 857
Qy 433 TATTGGAATCATAAACAAATTCAAAATCTTTGTTGCAATGTTGATGTTTGTAGTACAG 492
Db 858 GGAATAAAT---ACGACCAATATATATCGCTGCTGATCTATCTGATCTCTCTGAAAAGG 914
Qy 493 AAATAGATATAAATTAATTTATCAATTCAGCTCTATTATCTGTGGAAGATAGAGT 552
Db 915 ATTGTTATGATTTGAAACATAAATCTCGTTAAGGATATAAACAAGATGCTCTTACTCAG 974
Qy 553 TTTCAACAAAGGAAGTGGTCGACCACTGCTTTTATAAATCTACCAAGTGGATTTTGTAG 612

Db 975 AGTTAACTGCTTAACCTGTTGATCAAGCTTTTATTAACATATCCCATTAATTTGGTA 1034
Qy 613 TAAAGCAATTC-----GTTTCAGATTTTGGATTTCTTAGGTATTTGATGTAAAGCGTTG 663
Db 1035 AATACCAATACAGATATTAATGACATTTGAAGCAATGGAGTGTATGTCATGTTA 1094
Qy 664 ATATTAAATTCATTCGGTTTACTCGGTTGAAAAACAAAAAATCCAACTGCTAGTGACC 723
Db 1095 ATGTAGAGGCACTTTAGCTTTTGATAATATAGAGAAAAAGCAATCCAAACTTTTGAAGGAT 1154
Qy 724 ATAGCATTTGTAACCTTTTCCACAAATTTTATATAGCCTAGTCATATCATGATGAACACGAC 783
Db 1155 ATAGTGTATTACATATTTCTATGAATTTCTATATAATATAGTCACCTTATAGCAAAACGAT 1214
Qy 784 TTTTGGATATATCTCGGAGCGGTAGTCGGGTTTAAATTTATTTGGTATAGTCTCTATTTGT 843
Db 1215 TTTTGGATATCATGGTCTATTATAGGTTTGTCTCATATGTGGCATTGTGGCAATTTTTC 1274
Qy 844 TAGTTCCAAATTTTCGTAGAGATGGTGACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC 903
Db 1275 TAGTTCCGCAATCAGAAAAAGATGGTGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC 1334
Qy 904 AGAATGGACGCATATTTACATTTCTACAGTTTTCGATCGATGTATGTTGATGCTGAGGAC 963
Db 1335 GTAAATGGTAGGATTTTGTAGATTTCTATAAATTCAGATCAATCGAGTAGATGCGAACAAA 1394
Qy 964 GCAAAAAGACCTGCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTAAAAATGGGAA 1023
Db 1395 TTAAGAAAGATTTATTAGTTCACATCA-----ATGACGGGGCTAATGTTTAACTAGA 1449
Qy 1024 AAACGATCCTAGAATTTACTCCAATTTGGACATTTTCATACGCAAAAACAAGTTTACGAGT 1083
Db 1450 CGATGATCCTAGAATTTACTAAAATAGGAAAAATTTATTCG-AAAAACAAGCATAGATGAGT 1508
Qy 1084 TACCACAGTTTATAATGTTTAAATTTGCGATATGAGTCTAGTGGTACAGCTCCACCTA 1143
Db 1509 TGCCTCAATTTCTAATGTTTAAAGGTGATATGAGTTTATAGTAAAGACACGCCCTCCCA 1568
Qy 1144 CAGTTGATGAATTTGAAAAATATATCTCTGCTGCTCAAAAGAGAGAGTGTAGTTTAAACACAG 1203
Db 1569 CAGTTGATGAATATGAAGAAAGTATTAATTCACGCAAGACGCGCTTAGTTTAAAGCCAG 1628
Qy 1204 GGATTACAGGCTCTGCGCAGGTTAGTGGTCTAGTAAATATCAAGCTTCGACGACGTAG 1263
Db 1629 GAATCAGCTGTTTGTGCAAAATATCTGTAAGAAATAATTTACTGATTTGATGAAATCG 1688
Qy 1264 TCGGTTGGACCTTAGCATATGATTAATTTGCACTATCTGCTCAGATATTAATAATTTAT 1323
Db 1689 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTTATTTGGTTCAGATATTAAGATTTTC 1748
Qy 1324 TAAAGACAGTCGAAAGTTGCTATTGTTTGAGAGAGGGAAGTAAGTAAAGTA----- 1372
Db 1749 TCCTAACGCTAAAGTAGTTTTTACTCGGGAAGAGGAGCTAAGTAAAGGTAAGTTTGAAG 1808
Qy 1373 -----TATGAAAGTTTGTGTTGCTGGTTCTTTCAGGGGGGACATTTGACTCTGTTATTT 1426
Db 1809 GAATATAATGAATTTGTTCTGTTGTTTCAAGTGGTCTCAGCTAGCACACTTGAACT 1868
Qy 1427 GTTAAACCGTTTGGAAAGGAAGAACGTTTGGTGTAAACATTTGATTAAGAGATGTC 1486
Db 1869 TTTGAAACCCATTTGGGAAAAAGAGATAGGTTTGGGTAACTTTTGTGATAAGAGATGTC 1928
Qy 1487 AAGAAGTCTTTTGAAGATGAAGAAATGATTCATCTTCTTCAACAAATTCGCAATCT 1546
Db 1929 TAGGAGTATTTCTAAGAGAGAGAGATGTTATATCATTTGCTTCTTCCAAACAAACCGTAATGT 1988
Qy 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAAATTTTACGTGATGAGAAACGAGA 1606
Db 1989 CAAAAACCTTGGTAAAAAATATCTATTTCTAGCTTTTAAAGTCTTAGAAGAAAGAACGAGA 2048
Qy 1607 TGTTATTTATTTTATCTGCTGGCGGCTGCTGCCCTTCTTTTATCATCGGAAAACTATT 1666
Db 2049 TGTATCATATCATCTGCTGGCGCTGTAGCAGTACCACTTCTTTTATATTGGTAAAGTTATT 2108

| | | | |
|-----------------------------------------------------------|------|-----------------------------------------------------------------|------|
| Qy | 1667 | TGGACCAAGACGATTTATATTTGAAGTATTTTCATCGAGTTAATAAATCTACATTAAGTGG | 1726 |
| Db | 2109 | TGGGTGAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACCACTTTTGACAGG | 2168 |
| Qy | 1727 | AAAACTAGTTTATCCCGTAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGGT | 1786 |
| Db | 2169 | AAATATTAGTGTATCTGTAAACAGATAATTTATTTGTTTCAGTGGGAGAGAAATGAAGAAAGT | 2228 |
| Qy | 1787 | ATATCCTAAATCTATTAATCTTTGGGAGATTTTATTTTAATGATTTTGTAAACAGTAGGAAGT | 1846 |
| Db | 2229 | TTATCCTAAGGCAATTAATTTAGCAGGAATTTTATTAATGATTTTGTACACAGTAGGAGCA | 2288 |
| Qy | 1847 | CATGAACAACAGTTTATTCGATTGATATAAAGAGATTGATTTTATTTGAAGAAAAATGGAAGT | 1906 |
| Db | 2289 | CATGAACAACAGTTTCAACCGCTTATTATAAGAAAGTTGATAGATTAAAGGGACAGGTGCT | 2348 |
| Qy | 1907 | ATAACCGACGAAATATTTATTTCAACAGGATATCTGACTATATTTCCAGATAATTTGCAAG | 1966 |
| Db | 2349 | ATTGATCAAGAAGTGTTCATTTCAACCGGTTACTCAGACTTTGAACCTCAGAAATGTGAG | 2408 |
| Qy | 1967 | TATAAAAATTTCTCAGTTACAAAGAAATGGAACAATATATTAACAATCAGAAGTAGTT | 2026 |
| Db | 2409 | TGGTCAAAATTTCTCATATGATGATGAATGAATCTTACATGAAGAAGCTGAGATTGTT | 2468 |
| Qy | 2027 | ATTTGCCACGAGGCCCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAACAATTA | 2086 |
| Db | 2469 | ATCACACACGCGGTCACGCAACGTTTATGAATCGCGTTTCTAAAGGGAAAAAATATT | 2528 |
| Qy | 2087 | TTGTTTCTCAGACAAAAAAGTATGGTGAACATGTAAATGATCATCAAGTAGAGTTTGTA | 2146 |
| Db | 2529 | GTGGTTCTCAGACGAACAGTTTGGAGAGCATGTGAATATCATCAGGTGGATTTTTTG | 2588 |
| Qy | 2147 | AGAAGAATTTTCAAGAATAATAATATTTATTTATAGAAATATAGATGATTTGTTGAA | 2206 |
| Db | 2589 | AAAGAGTTATTCTTGAATATAGTTAGATTATATTTTGAATATCAGTGAATTAGAGAAAT | 2648 |
| Qy | 2207 | AAATATTAGTGAATTTCTAAGCAAC----TAACTTTACATCAATAATAATTTTTTTTGT | 2263 |
| Db | 2649 | ATTATTAGGAAAAAATATATCTACTAGTAAAGTAATATCACAAAACAATGATTTTTGT | 2708 |
| Qy | 2264 | GAAGAATTAACAAATAGTTTGAATAATTTTAAATGAGGATCAAGAAATGAATAATAAAA | 2323 |
| Db | 2709 | TCCTCTTTCAAAAATGAATTTCTTAACTATTTGAATAATAATATTTTGTGGAGAAAA | 2768 |
| Qy | 2324 | AGATGCA 2330 | |
| Db | 2769 | AATTGAA 2775 | |
| RESULT 21 | | | |
| AB050723 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| AB050723 | | | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| FEATURES | | | |
| source | | | |
| 1. .9987 | | | |
| /organism="Streptococcus agalactiae" | | | |
| /mol_type="genomic DNA" | | | |
| /db_xref="taxon:1311" | | | |
| 425. .430 | | | |
| 442. .1791 | | | |
| 442. .1791 | | | |
| /genes="cpsIB" | | | |
| /genes="cpsIB" | | | |
| /function="glucosyltransferase" | | | |
| /codon_start=1 | | | |
| /evidence=not experimental | | | |
| /transl_table=11 | | | |
| /product="CpsIB" | | | |
| /protein_id="BAB21597.1" | | | |
| /db_xref="GI:12697594" | | | |
| /translation="MIQTVVYFSASLTLTITPNFKSNKDLLFVLLIHIYIVFYLSDF | | | |
| YRDFSRGYLBEFPMVKYSFYIFISSLFFISKNSFTTTRLSFFTFIAMSILLYL | | | |
| LNSFLKYRYKYSKAFSRDVKVLIITNKDSLKMTFRNKYDHNIIYAVCIIDSSKDCY | | | |
| DLKHSRIINKDAL*SELTCLTVDQAFINIPITELFGYQIQDIINDIEMGVIVNV | | | |
| VEALSPNIGEKRIQTPEGYSVITYSMKFYSHLIAKRLDIMGALIGLLICGVAL | | | |
| FLVPIKDGGAIPFSONRVGRGRIPRFYKFSRMVDAEQIKDILVHNMOTGLMEK | | | |
| LDDPRITKIKGIRKTSIDELPQFINVLGMSLVGTREPTVDEYEKYNSTQRRLS | | | |
| FKPGITGLWISGRNNITDPDEIVKLDVQINWSIWSDKIILLTLKVLVLLGTGA" | | | |
| 1803. .1810 | | | |
| 1815. .2264 | | | |
| /genes="cpsIB" | | | |
| 1815. .2264 | | | |
| /genes="cpsIB" | | | |
| /codon_start=1 | | | |
| /evidence=not experimental | | | |
| /transl_table=11 | | | |
| /product="CpsIB" | | | |
| /protein_id="BAB21598.1" | | | |
| /db_xref="GI:12697595" | | | |
| /translation="MKICLVSSGGHIAHLNLLKPIWEKEDRFWVTFDKEDARSILRE | | | |
| EIVYHGFPTNRRNVKLAKVTLKRLERPDVITSSGAAVAPFPFYTGKLFCKT | | | |
| VYIEVDFRIDKPTLTGKLVYPVTDKFTVQWEEKVKVYPKAINLGGIF" | | | |
| 2249. .2256 | | | |
| /genes="cpsIB" | | | |
| 2284. .2746 | | | |
| /genes="cpsIB" | | | |
| 2284. .2746 | | | |
| /genes="cpsIB" | | | |
| /function="beta-1,4-galactosyltransferase" | | | |
| /codon_start=1 | | | |
| /evidence=not experimental | | | |
| /transl_table=11 | | | |
| /product="CpsIB" | | | |
| /protein_id="BAB21599.1" | | | |
| /db_xref="GI:12697596" | | | |
| /translation="MIFVTGTHEQQPNRLIKEVDRLKGTGAIDQEVFIQTGYSDREP | | | |
| QNCWMSKPLSDYDMNSTMKEAIVITHGGPATFMNAVSKGKTIVVPRQPGFGBHVN | | | |
| HQVDFLKLFLKVELDYILNISLENIIEKNISTSKVISQNNDFCSSFKNLSKLIFE | | | |
| " | | | |
| 2806. .2812 | | | |
| 2842. .3981 | | | |
| /genes="cpsIB" | | | |
| 2842. .3981 | | | |
| /genes="cpsIB" | | | |
| /function="putative CP polymerase" | | | |
| /codon_start=1 | | | |
| /evidence=not experimental | | | |
| /transl_table=11 | | | |
| /product="CpsIB" | | | |
| /protein_id="BAB21600.1" | | | |
| /db_xref="GI:12697597" | | | |
| /translation="MOLLALIVLLICSSYNEKMKFLNMAEIFFIVFVYVIVLVSIV | | | |
| LNSLFRSPFHRVIAAFNSLAVGVSLFVHYKNTNIELTKLKSFLFNAILPFLCG | | | |
| FLYYAIYFDVENVSIFGRNLIGSDWINGMHTQRAMAFPEYSNLIIPITITNIYIYI | | | |

(E-mail:miyake@proc.nubio.nagoya-u.ac.jp, Tel.81-52-789-4278,

Fax:81-52-789-3221)

Location/Qualifiers

1. .9987

/organism="Streptococcus agalactiae"

/mol_type="genomic DNA"

/db_xref="taxon:1311"

425. .430

442. .1791

/genes="cpsIB"

/genes="cpsIB"

/function="glucosyltransferase"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="CpsIB"

/protein_id="BAB21597.1"

/db_xref="GI:12697594"

/translation="MIQTVVYFSASLTLTITPNFKSNKDLLFVLLIHIYIVFYLSDF

YRDFSRGYLBEFPMVKYSFYIFISSLFFISKNSFTTTRLSFFTFIAMSILLYL

LNSFLKYRYKYSKAFSRDVKVLIITNKDSLKMTFRNKYDHNIIYAVCIIDSSKDCY

DLKHSRIINKDAL*SELTCLTVDQAFINIPITELFGYQIQDIINDIEMGVIVNV

VEALSPNIGEKRIQTPEGYSVITYSMKFYSHLIAKRLDIMGALIGLLICGVAL

FLVPIKDGGAIPFSONRVGRGRIPRFYKFSRMVDAEQIKDILVHNMOTGLMEK

LDDPRITKIKGIRKTSIDELPQFINVLGMSLVGTREPTVDEYEKYNSTQRRLS

FKPGITGLWISGRNNITDPDEIVKLDVQINWSIWSDKIILLTLKVLVLLGTGA"

1803. .1810

1815. .2264

/genes="cpsIB"

1815. .2264

/genes="cpsIB"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="CpsIB"

/protein_id="BAB21598.1"

/db_xref="GI:12697595"

/translation="MKICLVSSGGHIAHLNLLKPIWEKEDRFWVTFDKEDARSILRE

EIVYHGFPTNRRNVKLAKVTLKRLERPDVITSSGAAVAPFPFYTGKLFCKT

VYIEVDFRIDKPTLTGKLVYPVTDKFTVQWEEKVKVYPKAINLGGIF"

2249. .2256

/genes="cpsIB"

2284. .2746

/genes="cpsIB"

2284. .2746

/genes="cpsIB"

/function="beta-1,4-galactosyltransferase"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="CpsIB"

/protein_id="BAB21599.1"

/db_xref="GI:12697596"

/translation="MIFVTGTHEQQPNRLIKEVDRLKGTGAIDQEVFIQTGYSDREP

QNCWMSKPLSDYDMNSTMKEAIVITHGGPATFMNAVSKGKTIVVPRQPGFGBHVN

HQVDFLKLFLKVELDYILNISLENIIEKNISTSKVISQNNDFCSSFKNLSKLIFE

"

2806. .2812

2842. .3981

/genes="cpsIB"

2842. .3981

/genes="cpsIB"

/function="putative CP polymerase"

/codon_start=1

/evidence=not experimental

/transl_table=11

/product="CpsIB"

/protein_id="BAB21600.1"

/db_xref="GI:12697597"

/translation="MOLLALIVLLICSSYNEKMKFLNMAEIFFIVFVYVIVLVSIV

LNSLFRSPFHRVIAAFNSLAVGVSLFVHYKNTNIELTKLKSFLFNAILPFLCG

FLYYAIYFDVENVSIFGRNLIGSDWINGMHTQRAMAFPEYSNLIIPITITNIYIYI

| | | | |
|----|------|-------------------------------------------------------------------|------|
| Qy | 553 | TTTTCAACAAAGGAAGTGGTCGACCAACGCTCTTTATATAATCTACCAAGTGGTCTTTTACAGC | 612 |
| Db | 974 | AGTTAAACCTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCACTTGAATTTTGGTA | 1033 |
| Qy | 613 | TAAAGCAATTC-----GTTTCAGATTGTAGTTGGTTAGGTATTGATGTAAGCGTTG | 663 |
| Db | 1034 | AATACCAATACAGATATTTATTAATGACATTGAGCAATGGGAGTATGTTGTCATGTTA | 1093 |
| Qy | 664 | ATATTAATTCATTCGGTTTACTCGGTTGAAAAACAAAAATCCAACTGCTAGGTGACC | 723 |
| Db | 1094 | ATGTAGAGGCACTTAGCTTTGATATAATAGGAGAAAGCGAATCCAACTTTTGAAGGAT | 1153 |
| Qy | 724 | ATAGCATTTGAATCTTTTCCAAATTTTATAGCCTTAGTTCATATCATGATGAACGAC | 783 |
| Db | 1154 | ATAGTGTTATACATATTTCTATGAATTTCTATNAATATAGTACCTTATAGCAAAACGAT | 1213 |
| Qy | 784 | TTTTGGATATACTCGAGCGGTAGTCGGGTTTAAATTTTGTGGTATAGTTTCTATTTGT | 843 |
| Db | 1214 | TTTTGGATATCATGGGTGCTATTAATAGTTTGTCTCATATGTGGCATTTGGCAATTTTTC | 1273 |
| Qy | 844 | TAGTTCCAATTTATTCGTAGAGATGGTGCACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC | 903 |
| Db | 1274 | TAGTTCGCAATCAGAAAACGAGTGGACCGGCTATCTTTCTCAAAATAGAGTAGGTC | 1333 |
| Qy | 904 | AGATGGACGCATTTATCATTTCTACAAAGTTTCGATCGATGATGTTGTGCTGAGGAGC | 963 |
| Db | 1334 | GTAATGGTAGGATTTTATAGATTTCTATAAATTCAGATCAATCGCGAGTAGTCAGAACAA | 1393 |
| Qy | 964 | GCAAAAAGACTCTCTCAGCCAAAACAGATGCAAGGCTGGGTATGTTTAAATGGGAA | 1023 |
| Db | 1394 | TTAAGAAAGATTTATTTAGTTCACATCA-----AATGACGGGGCTAATGTTTAAAGTTAGA | 1448 |
| Qy | 1024 | AAACGATCCTAGAAATTTACTCCAATTTGGACATTTTCATACGCCAAAACAAAGTTTACGACAGT | 1083 |
| Db | 1449 | CGATGATCCTAGAAATTTACTAAAATAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT | 1507 |
| Qy | 1084 | TACCACAGTTTTATTAATGTTTAAATTTGCGGATATGAGTCTAGTTGGTACAGTCCACCTA | 1143 |
| Db | 1508 | TGCCTCAATTTCTAATGTTTTTAAAGGTGATATGAGTTTAGTAGGAACACGCGCTCCCA | 1567 |
| Qy | 1144 | CAGTTGATGAATTTGAAAAATATATCTCTGCTCAAAAGAGACGATTCGATTTTAAACGAG | 1203 |
| Db | 1568 | CAGTTGATGAATTTGAAAAATATATTTCAACGCAAGAGCGACGCTTAGTTTAAAGCCAG | 1627 |
| Qy | 1204 | GGATTACAGGCTCTCGCAGGTTAGTGGTCTGATGATATATACAGACTTCGACGACGTAG | 1263 |
| Db | 1628 | GAATCACTGGTTTGTGGCAATATCTGCTAGAAATATATTAATTAATTTGATGAAATCG | 1687 |
| Qy | 1264 | TTCCGTTGGACTTAGCATACATTTGATTAATTTGGACTATCTGCTAGATATTTAAATTTTAT | 1323 |
| Db | 1688 | TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTTGGTTCAGATATTAAGATTATTC | 1747 |
| Qy | 1324 | TAAAGACAGTCGAAAGTTGTTATTTGTGAGAGAGGGAAGTAAGTAAAGTA----- | 1372 |
| Db | 1748 | TCCTAACGCTTAAGGTAGTTTACTTCGGGACAGAGCTAAGTAAGGTAAAGTTTGAAG | 1807 |
| Qy | 1373 | -----TATGAAAGTTTGTGGTTCGTTCTTCAGGGGACATTTGACTCACCTTGTATTT | 1426 |
| Db | 1808 | GAATATAATGAATTTTGTCTGGTTGGTTCAAGTGGTGGTCAACCTAGCACACTTGAACCT | 1867 |
| Qy | 1427 | GTTAAACCGTTTGGAGGAGGAAGAACGTTTGGTGGTGAACATTTGATGAAGAGGATGC | 1486 |
| Db | 1868 | TTTGAACCCCATTTGGGAAAAAGAGATAGGTTTTTGGGTAACTTTTGTATGAAGAAGATGC | 1927 |
| Qy | 1487 | AAGAAGTCTTTTGAAGATGAAAAATGTATCCCATGTTACTTTCCCAAAATCCCAATCT | 1546 |
| Db | 1928 | TAGAGATATTTCTAGAGAGAGATTTGATATATCATTTCTTTTCCAAACACCCGTAATGT | 1987 |
| Qy | 1547 | CATTAAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA | 1606 |
| Db | 1988 | CAAAAACCTTGGTAAAAAATACTATCTTAGCTTTTAAAGTCTCTTAGAAAAAGAAACGACAGA | 2047 |

| | | | |
|----|------|------------------------------------------------------------------|------|
| Qy | 1607 | TGTTATTTATTTTCATCTGGCGCGTTCGTGTCCTCCCTCTTTTATCATCGGAAAACTATT | 1666 |
| Db | 2048 | TGTTATCATATCATCTGGTCCGCTGTAGCAGTACCAATCTCTTTTATATTTGGTAGTTATT | 2107 |
| Qy | 1667 | TGAGCAAAACGACGATTTATATTCGAAGTATTTGATCGAGTTTAAATAAACTCATATTAACTGG | 1726 |
| Db | 2108 | TGGCTGTAGACCGTTTATATAGGTTTTCGACAGATAGATAAACCAACTTTTGACAGG | 2167 |
| Qy | 1727 | AAAACTAGTTTATCCCGTAAACAGATATTTTTTATTTGTTTCAGTGGGAGAAATGAAGAAGGT | 1786 |
| Db | 2168 | AAAACTAGTGTATCTCTGTAACAGATAAATTTATTTGTTTCAGTGGGAGAAATGAANAAGT | 2227 |
| Qy | 1787 | ATATCTTAATCTTATTAACCTTGGGAGTATTTTTTATGATTTTGTAAACAGTAGGAACT | 1846 |
| Db | 2228 | TTATCTTAAGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTGCACAGTAGGAGCA | 2287 |
| Qy | 1847 | CATGAACCAACAGTTTAAATCGATTGATAAAGAGATTGATTATTTGAAAAAATAATGGAAGT | 1906 |
| Db | 2288 | CATGAACAGAGTTCAACCGTCTTATTAAGAAGTTTGTAGATTAAAGGGACAGGTGCT | 2347 |
| Qy | 1907 | ATAACCGACGAAATATTTATTTCAAAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG | 1966 |
| Db | 2348 | ATTGATCAAGAGTGTTCATTCAAAACGGTTTACTCAGACTTTTGAACCTCAGAAATTTGTCAG | 2407 |
| Qy | 1967 | TATAAAAAATTTCTCAGTTACAAAGAAATGGAACATATATTTAAACAATCAGAAGTAGTT | 2026 |
| Db | 2408 | TGCTCAAAATTTCTCTCATATGATGATATGAACCTTTTACATGAAGAAGCTGAGATTGTT | 2467 |
| Qy | 2027 | ATTGTCACGAGCGCCGCTACTTTTATGCAATTCATTATCCAAAGGAAAAACAATA | 2086 |
| Db | 2468 | ATCACACACGCGGTTCAGCAACGTTTATGAAATGCGAGTTCTTAAGGGAANAACAATTT | 2527 |
| Qy | 2087 | TTGTTTCTTAGACAAAAAAGTATGTTGAAACATGTAAATGATCATCAAGTAGAGTTGTA | 2146 |
| Db | 2528 | GTGGTTCTTAGACAAGAACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGAATTTTGT | 2587 |
| Qy | 2147 | AGAAGAAATTTTACAGATAATAATATTTTATTTATAGAAAATATAGATGATTTGTTGAA | 2206 |
| Db | 2588 | AAAAGAGTTATTTCTGAAATATGATGATTATATTTTGAATATCAGTGAATTAGAAGAT | 2647 |
| Qy | 2207 | AAAATATTCAAGTTCCTAAGCAAC--TAACTTTACATCAATAATAATTTTTTTTGT | 2263 |
| Db | 2648 | ATTATTAAGGAAAAAATATATCTACTAGTAAGATATATCAACAACAATGATTTTGT | 2707 |
| Qy | 2264 | GAAAGATTAACAAATAGTTGAAAAATTTAATCAGGATCAAGAAATGAATAATAAAA | 2323 |
| Db | 2708 | TCCTCTTTCAAAAATGAACCTTCTAACTATTTTGAATAAATAATATTTTGTGGAGAAAA | 2767 |
| Qy | 2324 | AGATGCA 2330 | |
| Db | 2768 | AATTGAA 2774 | |

RESULT 22

| | | | | |
|------------|-----------------------------------------------------------------------|-------------------------|--------|-----------------|
| AF332896 | 4768 bp | DNA | linear | BCT 07-JAN-2002 |
| LOCUS | Streptococcus agalactiae strain M781 | capsular polysaccharide | | |
| DEFINITION | bioynthesis gene cluster, partial sequence. | | | |
| ACCESSION | AF332896 | | | |
| VERSION | AF332896.1 | GI:18033333 | | |
| KEYWORDS | Streptococcus agalactiae | | | |
| SOURCE | Streptococcus agalactiae | | | |
| ORGANISM | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; | | | |
| REFERENCE | 1 (bases 1 to 4768) | | | |
| AUTHORS | Kong, P., Gowan, S., Martin, D., James, G. and Gilbert, G.L. | | | |
| TITLE | Serotype identification of group B streptococci by PCR and sequencing | | | |
| JOURNAL | J. Clin. Microbiol. 40 (1), 216-226 (2002) | | | |
| MEDLINE | 21635022 | | | |
| PUBMED | 11773119 | | | |
| REFERENCE | 2 (bases 1 to 4768) | | | |

| | |
|--------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| source | 1. .4768 /organism="Streptococcus agalactiae" /mol_type="genomic DNA" /strain="SG99/143" /serotype="III" /db_xref="taxon:1311" <1. .126 /gene="cpsD" /gene="cpsD" /codon_start=1 /transl_table=11 /product="CpsD" /protein_id="AAL56360.1" /db_xref="GI:18029140" /translation="AKEQMEQSGSKFLGIILNKVNESVATYGDYGNVGRDRKKR" 178. .1527 /gene="cpsE" 178. .1527 /gene="cpsE" /codon_start=1 /transl_table=11 /product="CpsE" /protein_id="AAL56361.1" /db_xref="GI:18029141" /translation="MIQTVVVFSGASLTLTITPNFKSNKDLLFVLLIHYIVFVLSDF YRDFWRCGLSEEFKWLKYSFYIIFISSLFFIPKNSFTTTRLSFTFIAMNSILLYL LNSFLKYRKYSAKFSRDTKVLITNKDSLKMTFRNKYDHNVIANCILDSSEKCY DLKNSLRINKDALISLTCITVDQAFINPIELFKYQIQDIINDIEMGVUNVN VEALSFNIGSKRIQTFGYSFYKSHLIARFLDITGAILIGLIGVAIL FLVQIRQDGPALFQSNVRNGRIIFRKYFRSMRVDAEQIKKDLVHNGMTGLMFK LEDDPRITKIGKIRKTSIDELPOFYNVLKGDMSLVGTRPPTVDEYKYNSTOKRRLS FKPGITGLWQISGRNNITDFDEIVKLDVQYINEMSIWSDIKIILLTLKVVLLGTGAK" 1551. .2000 /gene="cpsF" 1551. .2000 /gene="cpsF" /codon_start=1 /transl_table=11 /product="CpsF" /protein_id="AAL56362.1" /db_xref="GI:18029142" /translation="MKICLVGSGGHLAHLNLLKPIWEKEDRFWTFDKEDARSILRE RIVVHCRPTNRNVKNLKNITLAPKVLKERPDVIISGGAUAVPRFYIGKLGCKT VYIEVFRIDKPTLTGLKLVYPTDKFIQWHEMKVYPKAINLGGIP" 2000. .2473 /gene="cpsG" 2000. .2473 /gene="cpsG" /codon_start=1 /transl_table=11 /product="CpsG" /protein_id="AAL56363.1" /db_xref="GI:18029143" /translation="MIFVTGTHEQQFNRIKEVDRLKGTGAIODEVFIQGSDFEP QNCWSKELSDVNMVMEAEIVITHGPRATFMSVLSGLKLVVPRKQFGEHND HQIQFLKIAHLIPLWIEDVDGLAALKRNIAIEKYQGNNDMFCHLEKLEIGEI" 2470. .3615 /gene="cpsH" 2470. .3615 /gene="cpsH" /codon_start=1 /transl_table=11 /product="CpsH" /protein_id="AAL56364.1" /db_xref="GI:18029144" /translation="MRKYLDDYSLFYALWLVLVPMQYQFLIITIIVLLLLMKSEF RISNSNLSILFLMLFYILFALLINGTQEDITFPQFIABLLKLSTGPAHLPYNYRK ADFNSSVRNVKVVNVLVLIYLFPMKPLFGRELFSIEWFPHMRIRLAAYF EYATLIGQFIFLSPILFKPKQHMENILISLLTFCYSFGARILLVCMVLVASLL LDYILFKNLKLTFLGWTFLFITACFSNIWISIEKIMYRNQSTITRMIVYQ ESIIEVLGNILFGQIRIPSEGIIFLGSHTYISVFYRTSLIGLIVLFSAPILLYKE AISKNYKIYRFPFYLLCYTLPEEIDPNHWSIVLLFSTFGIVGRACK" |
| gene | 3612. .4580 /gene="cpsI" 3612. .4580 /gene="cpsI" /codon_start=1 /transl_table=11 /product="CpsI" /protein_id="AAL56365.1" /db_xref="GI:18029145" /translation="MKEKVTVIIPIYNSEAYLKECVOSVLOOQTHPLEVILIDGSTD NSGEICDNLQEDNRILVPHKNGGVSSARNIGLDKSTGREFITVDSDDFVAPNMIEI MLKNLITENADIAEVDFDISNERDYRKKRNFYVFNKNSLKEFLSGNVAENIVCT KLYKSIIGNLRFDENLIGEDLLFNCKLJCQEHRIVVDITSSITRYIKTSVWQK FNENSLDITIFNEISLVPARLANIYAEFLREKIKLRKMFELGNSIDNKIKVQRE IFFDKIDKSPFYKAVKYLKSLGLLSFLYLMKSPKLYMAYRRFQKQ" 4614. .>4768 /gene="cpsJ" 4614. .>4768 /gene="cpsJ" /codon_start=1 /transl_table=11 /product="CpsJ" /protein_id="AAL56366.1" /db_xref="GI:18029146" /translation="MIKKIEKDLISVIVPIYNVEDYLVECTIESLIVQYRNIIEILLIN DGSNDNCA" |
| CDS | ORIGIN Query Match 10.4%; Score 726.8; DB 1; Length 4768; Best Local Similarity 59.2%; Pred. No. 5.1e-73; Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6; QY 16 TGGCATTATTGATGATAGTAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCAA 75 DB 173 TAGCGATGATCAACAGTTGTAGTTATTTCTGCAAGTTTGACATTACATTAAATTA 232 QY 76 ATGTCGATTTAAATCG---TTCTGGAATTTTATCATAAATGATGGTTCATTATTTTGCAT 132 DB 233 CTCCCAATTTTAAAGCAATAAAGATTATTGTTGTTCTTATTGATACATTATATTTT 292 QY 133 TTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGTAATCTGATAGAGTTTGA 192 DB 293 TTTATCTTCTCGATTTTACAGAGACTTTTGAGTCTGCGGCTATCTTGAAGAGTTTAA 352 QY 193 AAACATTAATCTAGTATATATTTGCAATTTTCTTACGCGAGTATCATTTTGTGG 252 DB 353 TGGTATTGAAATACAGCTTTTACTATATTTTCATATCAAGTTTCATATTTTATTTT 412 QY 253 AGAATAATTTCCGACTTTTCAAGAGCGTGGTCCGTGTATTTTCCATTAATAAATTCGTTT 312 DB 413 AAAACTCTTTTACAACGACAGACTTTTCTTTTACITTTTATTGCTATGAATTCGATTT 472 QY 313 TGGTATACCTATTAAACGTAAATTATTAAGCAGTTTAAAGGATAGCTTTCTATTTTTCACA 372 DB 473 TATTATATCTATTGAAATTCATTTTAAATAATATTATCGAAAAATATTCTTACGCTAAAGTTT 532 QY 373 TCTATCAAAAAGACAGATCTTAATTAACCGCTGAACGATGGGAAATATGCAAGTTT 432 DB 533 CACGAGATACCAAAAGTTGTTTGATACGAATAAGGATCTTTATCAAAAATGACCTTTA 592 QY 433 TATTGAAATCACATAAAACAAATTTCAAAAAATCTTGTTCGATTTGGTAGTTTATAGGTACAG 492 DB 593 GGAATAAT---ACGACCATTAATTATATCGCTGCTGTATCTTGGACTCCTCTGAAAAGG 649 QY 493 AAATAGATAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 552 DB 650 ATTGTTATGATTGAAACATAAATCGTTAAGGATAATAAACAAGATGCTCTTACTTACG 709 QY 553 TTTCAACAGGGAAGTGTGCGACAGCTCTTTATATAAATCTACCAAGTGAAGTTTATAGACG 612 DB 710 AGTTAACCTGCTTAACCTGTTGATCAAGCTTTTATTAACATCCCAATGAAATTTATTTG 769 QY 613 TAAAGCAA-----TTTCGTTTTCAGATTTTGTAGTTGTTAGTTATGATGTAAGCGTTG 663 |


```
Db 770 AATACCAATACAAAGATATATTAATGACATGGAAGCAATGGGAGTGAATGTCATGTTA 829
Qy 664 ATATTAATTCANTCGGTTTACTCGGTGAAAAACAAAAATCAACCTGCTAGTGACC 723
Db 830 ATGTAGAGGCACTTAGCTTTGATATATAGGAGAAAGCGAATCCAAACTTTTGAAGAT 889
Qy 724 ATAGCATTTGAACCTTTTCCAAATTTTATATAGCCTAGTCATATCATGATGAACGAC 783
Db 890 ATAGTGTTATACATATCTATGAAATTTCTATAAATATAGTCACCTTATAGCAAAAGAT 949
Qy 784 TTTTGGATATACCTCGAGCGGTAGTCGGGTAAATTTATTTGGGTATAGTTTCTATTTTGT 843
Db 950 TTTTGGATATACCGGTGCTATATAGTTTGTCTATATGTTGGCAATTTGTGCAATTTTC 1009
Qy 844 TAGTTCOAATTAATTCGTAGAGATGGTGGACCGGCTATTTTGTTCAGAAACGAGTTGGAC 903
Db 1010 TAGTTCGCAATCAGAAAAAGATGGTGGACCGGCTATCTTTCTCAAAATAGAGTAGGTC 1069
Qy 904 AGAATGGACGATATTTACATTTCTACAAGTTTCAGTTCGATCGATGTATGTTGCTGAGGAC 963
Db 1070 GTAATGGTAGGATTTTGTAGATTTCTATAATTCAGATCAATCCGAGTAGTGCAGAACAA 1129
Qy 964 GCAAAAAGACTTCTCAGCCAAAACAGATGCAAGGTGGGTATGTTTAAAAATGGAA 1023
Db 1130 TTAGAAGATTTATAGTTTCAATCAATCAATGACAGG-----CTAATGTTAAGTTAGA 1184
Qy 1024 AAACGATCCTAGAAATTAATCTCAATTTGACATTTTCATAGCAAAACAAAGTTTACAGAGT 1083
Db 1185 AGATGATCCTAGAAATTAATCTCAATTTGACATTTTATTCG-AAAAACAAGCATAGATGAGT 1243
Qy 1084 TACCACAGTTTATAATGTTTAAATTCGCGATATAGTCTAGTTGGTACAGTCCACTA 1143
Db 1244 TGCCTCAATTTCTAATGTTTAAAGGTGATATGAGTTTAGTAGGAACACGCGCTCCCA 1303
Qy 1144 CAGTTGATGAATTTGAAAAATATATCTCTGCTCAAAAGAGACGATTCAGTTTAAACGAG 1203
Db 1304 CAGTTGATGATATGAAGAATTAATTTCAACGAGAGAGCGCTTAGTTTNAAGCAG 1363
Qy 1204 GGATTACAGGTCTCTGCGAGGTAGTGGTGTAGTATATACACAGCTTCGACGACGTAG 1263
Db 1364 GAATCACTGGTTTGTGCAAAATATCTGTAAGAAATATATTAATCTGATTTTGAATATCG 1423
Qy 1264 TTCGGTTGGACTTAGCATACATTAATTTGGACTATCTGGTCAGATTAATAATTTTAT 1323
Db 1424 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTGGTCAGATTAAGAATTATC 1483
Qy 1324 TAAAGACAGTGAAAGTTGTATTTGTTGAGAGAGGAAAGTAACTAAAGTA----- 1372
Db 1484 TCCTAACTAAAGGTAGTTTACTTCGGGACAGAGCTAGTAAGGTAAAGTTTGAAG 1543
Qy 1373 -----TATGAAGTTTGTGGTGGTCTTCAGGGGACATTTGACTCACTGTATTT 1426
Db 1544 GAATATAAGTAAATTTGCTGGTTGGTTCAAGTGGTGGTCTATAGTAAAGTTTGAACCT 1603
Qy 1427 GTTAAACCGTTTGGAGGAGAGAGACGTTTGGGTAAACATTTGATAAAGAGGATGC 1486
Db 1604 TTTGAACCCCATTTGGGAAAAGAGATAGGTTTGGGTAACTTTGATAAAGAGATGC 1663
Qy 1487 AAGAAGTCTTTTGAAGATGAAAAATGATCTCCATGTTTACTTCCAACAATCCCAATCT 1546
Db 1664 TAGAGTATTTCTAGAGAGAGATGTATATCATTTGCTCTTTCCAAACCAACCGTAATGT 1723
Qy 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTCAAAATTTTACGTGATGAGAAACAGA 1606
Db 1724 CAAAAACTTGTGTAATAATACTATTTCTAGCTTTTAAAGTCTCTTGAAGAAAGAACAGCA 1783
Qy 1607 TGTTATTTATTCATCTGGTGGCGGTGCTGCTCCCTTCTTTTACATCGGAAACHTATT 1666
Db 1784 TGTTATCATATCATCTGGTGGCGGTGAGCAGTACCAATCTTTTATATGTAAGTTATT 1843
Qy 1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTAATAAATCTACATTAACCTGG 1726
```

```
Db 1844 TGCTTCTAAGACCGTTTATATACAGGTTTTCGACAGATAGATAAAACCACTTTGCACAG 1903
Qy 1727 AAACTAGTTTATCCGTTAACAGATATTTTATTTGTTTCACTGGGAGAAATGAAAGAGGT 1786
Db 1904 AAAATTAGTGTATCTCTGTAAACAGATAAATTTATTTGTTCACTGGGAGAAATGAAAAAGT 1963
Qy 1787 ATATCTCTAAATCTATTAACTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAC 1846
Db 1964 TTATCTTAGGCAATTAATTTAGGAGAAATTTTAAAGTATTTTGTACAGTGGGAGCA 2023
Qy 1847 CATGAACAAACAGTTTAAATCGATTGATAAAGAGATTTGATTTATGAAAAAATGGAAGT 1906
Db 2024 CATGAACAGCAGTTCAACCGTCTTATTAAAGAGTTTGATAGATTAAAAGGACACAGGTGCT 2083
Qy 1907 ATAACCGACGAATATTTATTTCAACAGAGATATTTCTGACTATATTTCCAGATATTTGCAAG 1966
Db 2084 ATTGATCAAGAAGTGTTCATTCAACCGGTTTACTCAGACTTCGAACCTCAGAAATTTGTGAG 2143
Qy 1967 TATAAAAAATTTCTCAGTTTCAAGAAGAAATGGAACAATATATTAAACAAATCAGAGTAGTT 2026
Db 2144 TGGTCAAAATTTCTCTCATATGATGATATGAACTCTTTACATGAAGAAGAGCTGAGATTGTT 2203
Qy 2027 ATTGGACAGGAGCGCCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAAACAATTA 2086
Db 2204 ATCACACATGGCGCCGACGAGCTTTATGTGAGTTATTTCTTTAGGAAATTTACCAGTT 2263
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGTTGAACATGTAATGATCATCAAGTAGAGTTTGTGA 2146
Db 2264 GTTGTCTTAGGAGAAAGCAGTTTGGTGAACATATCAATGATCATCAAAATACAAATTTTGA 2323
Qy 2147 AGAAGAAATTTTACAAGATATAATATTTATTTATAGAAATATAGATGATTTGTTGAA 2206
Db 2324 AAAAAATTTGCCACCTGATCCCTGGCTTGGATTGAAGATGTAGATGAGACTTGGCGAA 2383
Qy 2207 AAAATTTTGAAGCTTTCTAAGCAAACTAACTTTACATCAAAATATATTTTGTGAA 2266
Db 2384 GCCTTGAAGGATATAGCTACAGAAATATACAGGAATATATGATATGTTTGTGAT 2443
Qy 2267 AGATTAACAAATAGTTGAAAAAT 2292
Db 2444 AAATTAGAAAAATATAGTGAAAT 2469

RESULT 24
AF363056 4768 bp DNA linear BCT 07-JAN-2002
LOCUS Streptococcus agalactiae strain GB00/031 capsular polysaccharide
DEFINITION biosynthesis gene cluster, partial sequence.
ACCESSION AF363056
VERSION AF363056.1 GI:18029147
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 4768)
AUTHORS Kong,F., Gowan,S., Martin,D., James,G. and Gilbert,G.L.
TITLE Serotype identification of group B streptococci by PCR and sequencing
JOURNAL J. Clin. Microbiol. 40 (1), 216-226 (2002)
MEDLINE 21635022
PUBMED 11773119
REFERENCE 2 (bases 1 to 4768)
AUTHORS Kong,F. and Gilbert,G.L.
TITLE Direct Submission
JOURNAL Submitted (22-WAR-2001) CIDMLS, ICPMR, Westmead Hospital, Darcy Road, Sydney, NSW 2145, Australia
FEATURES
source 1. 4768
/organism="Streptococcus agalactiae"
/mol_type="genomic DNA"
/strain="GB00/031"
/serotype="III"
```


| | | | | | | |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| gene | /db_xref="taxon:1311" <1..126 /gene="cpsd" <1..126 /gene="cpsd" /codon_start=1 /transl_table=11 /product="Cpsd" /protein_id="AAL56367.1" /db_xref="GI:18029148" /translation="AKEQMEQSGSKFLGIILNKYNESVATYDGYGNYGKRDRKRK" | gene | /transl_table=11 /product="Cpsi" /protein_id="AAL56372.1" /db_xref="GI:18029153" /translation="MKEKVTYIIPIYNSEAVLKBCVQSVOLOQTHPLIEVILDDGSTD NSGEICDNLSDENRILVPHKNGGVSSARNIGLXKSTGEFITFVDSDFVAPNMIEI MKNLITENADIARVDFDINSERDYRKKRRNFYKVPKNNLSKLEFLSGNRVENIVCT KLYKSIIGNLRFDPENLIGBDFLNCCKLQOEHRIVVDVTSSLYTYRIVKTSVNNQK FENSLDFPITIFNEISSILVPARLANYEAKLEKIKLCRKMFLGSGNIDNKIKVQRE IFFNDIKSPFYKAVKLSLKGLLSFYLMKSPKLYVMAYRRFQKQ" 4614..>4768 /gene="cpsu" 4614..>4768 /gene="cpsj" /codon_start=1 /transl_table=11 /product="Cpsj" /protein_id="AAL56373.1" /db_xref="GI:18029154" /translation="MIKKIEKDLISVIVPIYNVEDYLVECTIESLIVOTYRNIIEILLIN DGSTDNCA" | gene | 178..1527 /gene="cpsg" 178..1527 /gene="cpsg" /codon_start=1 /transl_table=11 /product="Cpsg" /protein_id="AAL56368.1" /db_xref="GI:18029149" /translation="MIQTVVVFVFSASLTLTITPNFKSNKDLLFVLLIHYIVFVLSDF YRDFWRSGLREPKMWLVKYSFYIIFISSLFFIKNSFTTTLSPFTFLMNSILLYL LNSFLKYRKYSAKFSRDTKVLITNKSLSKMTFRNKYDHNHYAVCILDSSEKCY DKNSURINIKDALTSELTCTVDQAFINPIELFGKYQIOTDINDIEAGLVVNVN VEALSFNIGIKRIQTFEGYSVITYSMKFKYSHLIAKRFIDITGAILIGICGIVAI FLVQIRKDGGAIPFSONVRNGRIFRFYKFRSMRVDAEQIKDILLVHNOQTMGLPK LEDDPRITKIGKIRKTSIDSLPOFYNVVLKGDMSIVGTRPPTVDYEKYNSTOKRRLS FKPGITGLWQISGRNNTDPEIVKLDVQYINWSIMSDIKIILLTLKVLVLGTCAK" 1551..2000 /gene="cpsf" 1551..2000 /gene="cpsf" /codon_start=1 /transl_table=11 /product="Cpsf" /protein_id="AAL56369.1" /db_xref="GI:18029150" /translation="MKICLVGSGGHLAHLNLLKPIWEKEDRFWVTFDKEDARSILRE EIVYHCPPTNENKVLVNTILAFKVLKRPDVLISSGAAVAVPPYIKGLPGCKT VYIEVDRIDKPTLTGKLVIPVTDKFIQVQEMKKVYPKAINLGLIF" 2000..2473 /gene="cpsg" 2000..2473 /gene="cpsg" /codon_start=1 /transl_table=11 /product="Cpsg" /protein_id="AAL56370.1" /db_xref="GI:18029151" /translation="MIFVTVGTHQEQFNRLIKEVDRLKGTGCAIDQEVFIQYGSDFEP QNCWSKPLSYDDMNSYNKEABIVITHGGPATFMSVISLGLPVPVVPKRPQFGEHND HQIQFLKIAHLYPLAWIEDVDGLABALKRNIATEKYQGNNDMFCHEKLEKIGSI" 2470..3615 /gene="cpsH" 2470..3615 /gene="cpsH" /codon_start=1 /transl_table=11 /product="CpsH" /protein_id="AAL56371.1" /db_xref="GI:18029152" /translation="MRKYLDLDYSLFYALWVLIVPNQWYOFLLIITIIVLLLMKSEF RISNSISILMLFYLFALIRGTQEDITPORAELLLKISTGYALFFYNYRK ADFNSSVNVNKKVYFLVLTIVLYLFPMLKPTLFGRELFSIEWFPHMRILAAVYR BYATLIGQFILFSYLPFLPKQKMENTILISLLTICYSYFSGARILIVCMVLVASLL LDYILKTNLKLFTGMLTFLFITACFSYNIWSIEKIMYRQSTITRMIVQ ESLIEVLKGNLFGGKIPSGEFLGSHSTYSVFTYRSLGLVLFYSFALILYKE AISKNYKIYRUFFTYLLCYTLFEEIDPNHNSIVLLFSTFGIVGRKK" 3612..4580 /gene="cpsI" 3612..4580 /gene="cpsI" /codon_start=1 | Query Match 10.4%; Score 726.8; DB 1; Length 4768; Best Local Similarity 59.2%; Pred. No. 5.1e-73; Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6; Qy 16 TGGCATTAATTGATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACCAA 75 Db 173 TAGCGATGATTCAACACAGTTGTAGTTTATTTTCTGCAAGTTTGACATTAACATTA 232 Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATAATGATGTTTCATTATTTTGCAT 132 Db 233 CTCCAAATTTAAAGCAATAAGATTTATTGTTGTTCTATTGATACATTAATTGTTT 292 Qy 133 TTTTATATCTCGTATGCGAGTTGAAATTTAGTATAGAGTAATCTGATAGAGTTTGAAA 192 Db 293 TTTATCTTCTCGATTTTACAGAGACTTTTCGAGTCGTGGCTATCTTGAAGAGTTTAAAA 352 Qy 193 AAACATTAACTAGTATATAATATTGCAATTTTCTTACGGCAGTATCATTTTGTGGTGG 252 Db 353 TGGTATTGAATACAGCTTTTACTATATTTTATATCAATCAAGTTTCATTAATTTATTTTA 412 Qy 253 AGAATAATTTCCGACTTTCAAGACGTGGTGGCGGTATTTTACATTAATTAACCTTCGTTT 312 Db 413 AAAAATCTTTTACAAACGACGACTTTCCTTTTATTCTATGCTATGAATTCGATTT 472 Qy 313 TGGTATACCTATTAAACGTAATTAAGCAGTTTAAGGATAGCTTTCTATTTCGACAA 372 Db 473 TATTATATCTATTGAATTCATTTTAAATATTATCGAAATATTCTTACGCTAAGTTT 532 Qy 373 TCTATCAAAAAAAGACGATTCATAATTACACGGCTGAACGATGGGAAATATGCAAGTTT 432 Db 533 CACGAGATACCAAAGTTGTTTGTATACGAATAAGGATTCCTTATCAAAAATGACCTTTA 592 Qy 433 TATTGTAATCACATAAACAATTCAAAAAATCTTGTGTCATTTGTTAGTTTGTAGTACAG 492 Db 593 GGAATAAAT---ACGACCATTAATTATATCGCTGTCTGTATCTTGACCTCTCTGAAAAGG 649 Qy 493 AAATAGATAAAATTAATTTATCATTTACCGCTCTATTATTCTGTGGAAGAAGCTATAGAGT 552 Db 650 ATTGTTATGATTGAAACATACATCTCGTTAAGGATAATAACAAGATGCTCTTACTTCAG 709 Qy 553 TTTCAACAAGGAAGTGGTCCACACGCTCTTTTATAAATCTACCAAGTCAGTTTTTAGACG 612 Db 710 AGTTAACTGCTTAACGTGTCATCAAGCTTTTATTAAACATACCCATTGAATTTTGGTA 769 Qy 613 TAAAGCAA-----TTCGTTTCAGATTTTGTAGTTTGTAGTATTGATGTAAAGCGTTG 663 Db 770 AATACCAATACAGATATTATTAAATGACATTGAAGCAATGGAGCTGATTGTCATGTTA 829 Qy 664 ATATTAAATTCATTGGTTTCTGCTGTTTGAACCAAAAAATCCAATCTGCTAGTGACC 723 |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

| | | | |
|----|------|------------------------------------------------------------------|------|
| Ds | 830 | ATGTAGAGGCACTTAGCTTTTGATATATAGGAGAAAGCGAATCCAAACCTTTTGAAGGAT | 889 |
| Qy | 724 | ATAGCATTTGTAACCTTTTCCAAAATTTTATAGCCCTAGTCATATCATGATGAAGAACGAC | 783 |
| Ds | 890 | ATAGTGTATTACATATCTATGAAATTTCTATAAATATAGTCACCTTTATAGCAAAACGAT | 949 |
| Qy | 784 | TTTTGGATATACTCGGAGCGGTAGTCGGGTTAAATTTTCTGGGTATAGTTTCTATTTTGT | 843 |
| Ds | 950 | TTTTGGATATACAGGGTGCTATTATATAGTTTGTCTATATGGAATTTGTGGCAATTTTTC | 1009 |
| Qy | 844 | TAGTTCCAAATTTATTCGTAGAGATCGTGCACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC | 903 |
| Ds | 1010 | TAGTTCCGCAATCAGAAAAGATGCTGCACCGGCTATCTTTCTCAAAATAGAGTAGGTC | 1069 |
| Qy | 904 | AGAATGAGCGATATTTACATTTACAAAGTTTCGATCGATGTATGTTGATCGTGGAGGC | 963 |
| Ds | 1070 | GTAATGCTAGGATTTTATAGATTTCTATAAATTCAGATCAATGCGAGTAGATGCGAACAAA | 1129 |
| Qy | 964 | GCAAAAAGACTTCTCAGCCCAAAACAGATGCAAGGTTGGGTATGTTTAAATGGAA | 1023 |
| Ds | 1130 | TTAAGAAAGATTTTATAGTTCACAAATCAAAATGACAGGG-----CTAATGTTTAAAGTTAGA | 1184 |
| Qy | 1024 | AAACGATCCTAGATTTACTCCAAATGGACATTTTCATACGCAAAAACAAAGTTTACACGAGT | 1083 |
| Ds | 1185 | AGATGATCCTAGATTTACTAAAATAGGAAATTTTATTCG-AAAAACAAGCATAGATGAGT | 1243 |
| Qy | 1084 | TACACAGTTTATAATGTTTAAATGCGGATATGAGTCTAGTTGTTGATACAGTCCACTA | 1143 |
| Ds | 1244 | TGCTCAATTTCTATAATGTTTAAAGGTGATATGAGTTTAGTAGGAACACGCGCTCCCA | 1303 |
| Qy | 1144 | CAGTTGATGATTTGAAAATATACTCTGCTCAAAAGACGATTCAGTTTAAACAG | 1203 |
| Ds | 1304 | CAGTTGATGATATGAAAAGTATTAATTCACGCAAGACGCGCTTAGTTTAAAGCCAG | 1363 |
| Qy | 1204 | GGATTACAGGCTCTCGCAGGTTAGTGTCTGCTAGTAATATCACAGACTTCGACGACGTAG | 1263 |
| Ds | 1364 | GAATCATTGTTTGTGCAATATCTGGTAGAATATATTTACTGATTTTGTAGMAATCG | 1423 |
| Qy | 1264 | TTGCGTTGGACTTAGCATACATTCGATTAATTTGGACTATCTGCTCAGATATTTAAATTTAT | 1323 |
| Ds | 1424 | TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTAAGATTATTC | 1483 |
| Qy | 1324 | TAAAGACAGTCAAAAGTTGATTTGTTAGAGAGGGAAGTAAGTAAAGTA----- | 1372 |
| Ds | 1484 | TCCTAAACATAAAGGTAGTTTACTCGGGAACAGGACTAAAGTAAAGGTGTTGAAG | 1543 |
| Qy | 1373 | -----TATGAAAGTTTGTGTTGGTTCCTTCAGGGGACATTTGACTCACTTGATTTT | 1426 |
| Ds | 1544 | GAATATATGAAATTTGTCTGGTTGGTTCAAGTGGTGGTCACTAGCACACTTGAACCT | 1603 |
| Qy | 1427 | GTTAAACCGTTTGGGAAGGAAGAACGTTTTTGGGTAAACATTTGATAAAGAGGATGC | 1486 |
| Ds | 1604 | TTTGAACCCATTTGGGAAGGAAGATAGGTTTGGGTAACTTTGATAAAGAGATGC | 1663 |
| Qy | 1487 | AAGAGTCTTTTGAAGATGAAAAATGTATCCATGTTACTTCCACAAATCGCATCT | 1546 |
| Ds | 1664 | TAGGATATTTCTAAGAGAAGAGATTTGATATCATTTGCTTCTTCTTCCAAACCCGTAATGT | 1723 |
| Qy | 1547 | CATTAAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACAGA | 1606 |
| Ds | 1724 | CAAAAATTTGGTAAAAATCTATTTCTAGCTTTTAAAGTCTTTAGAAAAGAGAACCCAGA | 1783 |
| Qy | 1607 | TGTTATTTATTCATCTGGTGGCGGCTGTGCTGCCCTTCTTTTACATCGGAAACATTT | 1666 |
| Ds | 1784 | TGTTATCATATCATCTGGTGGCGCTGTAGCAGTACCATTCTTTTATTTGTAAGTTATT | 1843 |
| Qy | 1667 | TGGAGCAAGACGATTTATTTAGATATTTTGAATGTCAGTTTAAATTAATCTAATTAACG | 1726 |
| Ds | 1844 | TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACCACTTTTGAAG | 1903 |
| Qy | 1727 | AAAACTAGTTTATCCCGTAACAGATATTTTATTTGTTTTCAGTGGGAAGAAATGAAGAAGT | 1786 |
| Ds | 1904 | AAATTTAGTGTATCTCTGTAAACAGATAATTTTATTTGTTTTCAGTGGGAAGAAATGAAGAAGT | 1963 |

| | | | |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|--------|
| Qy | 1787 | ATATCTTAATCTATTAACTTTGGGAGTATTTTTTAAATGATTTTTTTGTAACAGTAGGAAC | 1846 |
| Ds | 1964 | TTATCTCAAGGCAATTAATTTAGGAGGAATTTTTTAAATGATTTTTTTGTCACAGTGGGACA | 2023 |
| Qy | 1847 | CATGAACAACAGTTTAAATCGATTGATAAAGAGATTTGATTTATTTGAAAAAAATGGAAGT | 1906 |
| Ds | 2024 | CATGAACAGCGATTCAACCGCTCTTATTAAGAAGATTGATAGATTAAAGGACAGGTGCT | 2083 |
| Qy | 1907 | ATAACCGACGAAATATTTATTTCAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG | 1966 |
| Ds | 2084 | ATTGATCAAGAAGTGTTCATTCAACCGGTTACTCGACTTTCGAACCTCGAATTTGTCAG | 2143 |
| Qy | 1967 | TATAAAAATTTCTAGTTTCAAGAAGAAATGGAACAATATATTAAACAATCAGAAGTAGTT | 2026 |
| Ds | 2144 | TGCTCAAAATTTCTCTCATATGATGATGAACTTTTACATGAAGAAGAGCTGAGATTGTT | 2203 |
| Qy | 2027 | ATTGTCACGAGGCGCGCTACTTTTATGAAATTTTATCCAAAGGAAAAACAATTA | 2086 |
| Ds | 2204 | ATCACACATGGCGCCGACGCGTTTATGTCAGTTTATTTCTTTAGGGAATTTACCAGTT | 2263 |
| Qy | 2087 | TTGTTTCTTAGACAAAAAAGTATGTTGAACATGTAAATGATCATCAAGTAGAGTTTGT | 2146 |
| Ds | 2264 | GTTGTTCTTAGGAGAAAGCAGTTTGGTGAACATATCAATGATCATCAATACAAATTTTA | 2323 |
| Qy | 2147 | AGAAGAAATTTTCAAGATAATAATATTTTATTTATAGAAAATATAGATGATTTGTTGAA | 2206 |
| Ds | 2324 | AAAAAAATTTGCCACCTGTATCCCTTGGCTTGGATTGAAGATGTAGATGACTTTCGCGAA | 2383 |
| Qy | 2207 | AAAATTTATGAAGTTTCTAAGCAAACTAATCTTTACATCAATTAATATTTTGTGTA | 2266 |
| Ds | 2384 | GCCTTGAAGAAGAAATATAGCTACAGAAAATATCAGGAAATATATGATGATTTTGTGAT | 2443 |
| Qy | 2267 | AGATTAACAAATAGTTGAAAAATTT | 2292 |
| Ds | 2444 | AAATTAGAAAATTTAGTGGAAT | 2469 |
| RESULT 25 | | | |
| AF163833 | | | |
| LOCUS | 17276 bp | DNA | linear |
| DEFINITION | Streptococcus agalactiae capsular polysaccharide synthesis locus, complete sequence. | | |
| ACCESSION | AF163833 | L09116 | |
| VERSION | AF163833.1 | GI:5823209 | |
| KEYWORDS | Streptococcus agalactiae | | |
| SOURCE | Streptococcus agalactiae | | |
| ORGANISM | Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus. | | |
| REFERENCE | 1 | (bases 4755 to 6119) | |
| AUTHORS | Rubens,C.E., Heggen,L.M., Haft,R.F. and Wessels,M.R. | | |
| TITLE | Identification of cpsB, a gene essential for type III capsule expression in group B streptococci | | |
| JOURNAL | Mol. Microbiol. 8 (5), 843-855 (1993) | | |
| MEDLINE | 93360815 | | |
| PUBMED | 8355611 | | |
| REMARK | cpsB now called cpsE | | |
| REFERENCE | 2 | (bases 15471 to 16670) | |
| AUTHORS | Haft,R.F., Wessels,M.R., Mebane,M.F., Conaty,N. and Rubens,C.E. | | |
| TITLE | Characterization of cpsF and its product CMP-N-acetylneuraminic acid synthetase, a group B streptococcal enzyme that can function in K1 capsular polysaccharide biosynthesis in Escherichia coli | | |
| JOURNAL | Mol. Microbiol. 19 (3), 555-563 (1996) | | |
| MEDLINE | 96228704 | | |
| PUBMED | 8830246 | | |
| REMARK | cpsF now called neuA | | |
| REFERENCE | 3 | (bases 1 to 17276) | |
| AUTHORS | Chaffin,D.O., Beres,S.B., Yim,H.H. and Rubens,C.E. | | |
| TITLE | The serotype of type Ia and III group B streptococci is determined by the polymerase gene within the polycistronic capsule operon | | |
| JOURNAL | J. Bacteriol. 182 (16), 4466-4477 (2000) | | |
| MEDLINE | 20372630 | | |

| | | |
|-----------|--------------------------------------------------------------------|---------------|
| 10913080 | 2570. .2576 | RBS |
| REFERENCE | /gene="cpsB" | |
| AUTHORS | 2586. .3317 | CDS |
| TITLE | /gene="cpsB" | |
| JOURNAL | /function="tyrosine phosphatase" | |
| | /codon_start=1 | |
| | /transl_table=11 | |
| | /product="CpsB" | |
| | /protein_id="AAB00361.1" | |
| | /db_xref="GI:153590" | |
| | /translation="MIDIHSHIVFDVDDGPKTLEESLSTEEVSROGVRIIVSTSHR | |
| | KGMETPEDIIKKNFSIVKHETEKPEHLOILYGGELVYTSMDLEKILKLOIPTLANT | |
| | KPALIEFSMOTSQWDIHTALSIVMLGITVVAHIERYNALENQKERVKEIINMGCT | |
| | QINSGFHILKQIFDNDRKFRKPRKARYPLEENLVHFVASDMHNDVDRPPFLAEAYKILC | |
| | RDQKGRANQLFIENAQSILKNHYI | |
| | 3316. .4018 | gene |
| | /gene="cpsC" | |
| | 3316. .3321 | RBS |
| | /gene="cpsC" | |
| | 3326. .4018 | CDS |
| | /gene="cpsC" | |
| | /note="putative polysaccharide chain length regulator" | |
| | /codon_start=1 | |
| | /transl_table=11 | |
| | /product="CpsC" | |
| | /protein_id="AAB00362.1" | |
| | /db_xref="GI:153591" | |
| | /translation="MNKIANTVEINIFNLKKLWKKKFLITFEVAFATAGLIFYSLF | |
| | IVTQYTSSTRIYVINPNTNNSITADQLQAGSLFANDYKEIITSVDLEKVISSSKL | |
| | NYPSSQLQKITVTSILKDXTRVISISVEDANPKMSQKLANSVREAAVSKIKAVTQVEDI | |
| | TTELEGNLPKAPSPFNKNVNLGIVAGLSTIVLIMGILDDRVTNTEDEIEKVLGI | |
| | TSLGIVDPDLNKL" | |
| | 4018. .4718 | gene |
| | /gene="cpsD" | |
| | 4018. .4024 | RBS |
| | /gene="cpsD" | |
| | 4029. .4718 | CDS |
| | /gene="cpsD" | |
| | /function="autophosphorylating tyrosine kinase" | |
| | /note="CPS chain length regulator/exporter" | |
| | /codon_start=1 | |
| | /transl_table=11 | |
| | /product="CpsD" | |
| | /protein_id="AAB00363.1" | |
| | /db_xref="GI:1323730" | |
| | /translation="WTRLEIVDSKLRQAKTEEVFAIRTNIOFSGKENKILAIASVR | |
| | EGEKSITSTSLASLAQAGFKTLIDADTRNSVMGTFKATGTIKGLTNYLSGNADL | |
| | GDIICETNVPRLMVVPSGKVPFPNTALLQNAVFNKMEIAKNIPDIYIIDTTPPIGLV | |
| | DAALIANACDGFILVTQAGRIKKNYVEKAQEQSGKFLGIILKNVNESVATYGIQ | |
| | GNYGRDRKXK" | |
| | 4755. .6119 | gene |
| | /gene="cpsE" | |
| | 4755. .4761 | RBS |
| | /gene="cpsE" | |
| | 4770. .6119 | CDS |
| | /gene="cpsE" | |
| | /function="glucosyl-1-phosphate transferase" | |
| | /note="glucosyl transferase; previously described as CpsD" | |
| | /citation=[1] | |
| | /codon_start=1 | |
| | /transl_table=11 | |
| | /product="CpsE" | |
| | /protein_id="AADS3066.1" | |
| | /db_xref="GI:5823212" | |
| | /translation="MIOTVVYFVSASLTLTITPNPKSNKDLLFVLLIHYIVFLSDP | |
| | YRDFWSRGYLEEFKQWLVKSYFYFIISSSLFFIPKNSFTTSLRSTFTTAMNSILLYL | |
| | INSLFTKRYKSYAKFSRDTKVLITNKLSLSPKNTFRNKYDHYIIVACILDSSEKOCY | |
| | DLKNSLRINKDALSTELTCLTVDOAFINPIELFKQYQIDIIEMAMGIVVNV | |
| | VEALSFNICGKRIQTPREGSVITYSMKFKYSHLIAKRLDTGAIIGLIGIIVAI | |
| | FLVQIRKGGPAILFQSNRVGRNGRIFRFYKFRSMRVDAEQIKKOLLVENQMTGLMFK | |
| | LEDPPRIKTKGGRIRKTSIDELPOFINVLGDSMSLAGTRPTTVEYKNTQKRRLS | |
| | FKPGITGLMQISGRNNITDFDEIVKLVDSINWSWSDIKIILLTLKVLLTGTGAK" | |
| | 6131. .6592 | gene |
| | /gene="cpsB" | |
| 10913080 | 2570. .2580 | gene |
| REFERENCE | 4 (bases 2196 to 5903) | |
| AUTHORS | Rubens,C.E. | |
| TITLE | Direct Submission | |
| JOURNAL | Submitted (27-JAN-1993) Division of Infectious Disease, Immunology | |
| | and Rheumatology, Children's Hospital and Regional Medical Center, | |
| | 4800 Sand Point Way N.E., Seattle, WA 98105, USA | |
| | 5 (bases 4495 to 6194) | |
| | Yim,H.H. | |
| REFERENCE | Direct Submission | |
| AUTHORS | Submitted (16-MAY-1996) Division of Infectious Disease, Immunology | |
| JOURNAL | and Rheumatology, Children's Hospital and Regional Medical Center, | |
| | 4800 Sand Point Way N.E., Seattle, WA 98105, USA | |
| | 6 (bases 1 to 17276) | |
| | Chaffin,D.O., Yim,H.H., Beres,S.B., Sweet,E.S., Nittayajarn,A. and | |
| | Rubens,C.E. | |
| | Direct Submission | |
| | Submitted (29-JUN-1999) Division of Infectious Disease, Immunology | |
| | and Rheumatology, Children's Hospital and Regional Medical Center, | |
| | 4800 Sand Point Way N.E., Seattle, WA 98105, USA | |
| | On Sep 5, 1999 this sequence version replaced gi:1323729. | |
| COMMENT | Location/Qualifiers | |
| FEATURES | 1. .17276 | |
| source | /organism="Streptococcus agalactiae" | |
| | /mol_type="genomic DNA" | |
| | /strain="COH1" | |
| | /serotype="III" | |
| | /db_xref="taxon:1311" | |
| | complement(7. .930) | |
| gene | /gene="cpsY" | |
| CDS | /gene="cpsY" | |
| | /note="LySR homolog; putative transcriptional regulator" | |
| | /codon_start=1 | |
| | /transl_table=11 | |
| | /product="CpsY" | |
| | /protein_id="AADS3064.1" | |
| | /db_xref="GI:5823210" | |
| | /translation="MRIQLOQYIKIVETGSMNEAAKQLYITOPISLNAVRLNLETEMG | |
| | IQIFRNPKGITLTKDGMETLSYARQILEOTALLAEERYKGDNTSRELFVSSSHAFV | |
| | YNAFVALNGMDTQYELFLRETTWEIIDDVKNFRSEIGVLFNLSYNRDLTKLFDD | |
| | NSLIATITTFTHPFSKSNPLANRKLKMKOLEDYPILSYDQGLNLSYFSPSEWMS | |
| | QIPHPKSIWSDRATFLNMLTGLDGYTVATGILNSKLNGLDEIVAIPLDVDDVIDVIYI | |
| | RHDKANLSMGQKFTDYLLREVSFNHNKQS" | |
| | 932. .951 | repeat region |
| | 1047. .1052 | -35 signal |
| | 1070. .1075 | -10 signal |
| | 1083 | misc_feature |
| | /note="CPSIII operon transcriptional start site" | |
| | /evidence=experimental | |
| gene | 1108. .2580 | |
| | /gene="cpsA" | |
| RBS | 1108. .1114 | |
| | /gene="cpsA" | |
| CDS | 1120. .2580 | |
| | /gene="cpsA" | |
| | /note="LySR homolog" | |
| | /codon_start=1 | |
| | /transl_table=11 | |
| | /product="CpsA" | |
| | /protein_id="AADS3065.1" | |
| | / | |

| | | | |
|----|------|-----------------------------------------------------------------|------|
| QY | 1787 | ATATCCTAAATCTATTAACTTCGGGAGATATTTTTTAATGATTTTTTGTAAACAGTAGGAACT | 1846 |
| Db | 6556 | TTATCCTAAGGCAATTAATTTAGGAGGAATTTTTTAATGATTTTTTGTACAGTGGGACA | 6615 |
| QY | 1847 | CATGAACAACAGTTTAAATCGATTGATAAAGAGATTGATTATTCGAAAATAATGGAAGT | 1906 |
| Db | 6616 | CATGAACAGCAGTTCAACCGTCTTATTAAGAGAGTTGATAGATTAAAGGGACAGTGCT | 6675 |
| QY | 1907 | ATAACCGACGAATATTTATTCAAAACAGGATATTTCTGACTATATTTCCAGAATATTTGCAAG | 1966 |
| Db | 6676 | ATTGATCAAGAGTGTCTATTCAACCGGTTACTCAGACTTCGAACCTCAGAAATTTGTCAG | 6735 |
| QY | 1967 | TATAAAAAATTTCTCAGTTACAAAAGAAATGGACAAATATATTAACAATCAGAACTAGTT | 2026 |
| Db | 6736 | TGGTCAAAAATTTCTCTCATATGATGATATGAACTCTTTACATGAAAAGAGCTGAGATTGTT | 6795 |
| QY | 2027 | ATTTGCCACGAGGCCCGCTACTTTTATGAATTCATTATCCAAAAGGAAAACAAATTA | 2086 |
| Db | 6796 | ATCACACATGGCGGCCCGACGTTTATGTCAAGTTATTTTAGGGAAATTACCGTT | 6855 |
| QY | 2087 | TTGTTTCTTAGACAAAAAAGATGTTGTAACATGTAATGATCATCAAGTAGAGTTTGTA | 2146 |
| Db | 6856 | GTGTTCTTAGGAGAAGCAGTTTGGTGAACATATCAATGATCATCAATACAATTTTA | 6915 |
| QY | 2147 | AGAAGAAATTTACAAGATAATAATTTTATTTATAGAAAAATATAGATGATTTGTTGAA | 2206 |
| Db | 6916 | AAAAAAATTTGCCACCTGTATCCCTTGGCTTGGATTGAAGATGTAGATGGACTTCGGAA | 6975 |
| QY | 2207 | AAAAATTATTGAAGTTTCTAAGCAACTAATTTTACATCAAAATAAATTTTTTTGTGAA | 2266 |
| Db | 6976 | GGTTGAAAAGGAATATAGCTACAGAAAATATCAGGGAATAATGATATGTTTGTGAT | 7035 |
| QY | 2267 | AGATTAAAAACAATAGTTGAAAAAT | 2292 |
| Db | 7036 | AAATTAGAAAAAATTTAGTGAAAT | 7061 |

Search completed: April 29, 2005, 10:37:38
Job time : 28940 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2005, 15:28:57 ; Search time 3184 Seconds
(without alignments)
12999.633 Million cell updates/sec

Title: US-09-767-041-9

Perfect score: 6992
Sequence: 1 atcgccaaacgaattggca.....gcaaacagcaaaagaagctt 6992

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004as:*
- 13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|---------------------|
| 1 | 6992 | 100.0 | 6992 | 3 | Aaz60930 Nucleotid |
| 2 | 1195.8 | 17.1 | 26281 | 3 | Aaz60929 Nucleotid |
| 3 | 955.6 | 13.7 | 17468 | 6 | ABA01441 Streptoco |
| 4 | 889 | 12.7 | 16032 | 13 | Adsl3200 S. thermo |
| 5 | 798.8 | 11.4 | 110000 | 6 | Continuation (12 o |
| 6 | 727.8 | 10.4 | 6865 | 6 | ABK90550 Betal, 3-g |
| 7 | 726.8 | 10.4 | 17276 | 8 | ACA64723 Streptoco |
| 8 | 726.8 | 10.4 | 17276 | 10 | ADF43363 Streptoco |
| 9 | 721.4 | 10.3 | 25020 | 11 | ADO40235 S. agalac |
| 10 | 649.2 | 9.3 | 2226 | 12 | Adm79774 Group B S |
| 11 | 428.4 | 6.1 | 6850 | 4 | Aah43197 Portion o |
| 12 | 423.2 | 6.1 | 1368 | 4 | Aas55950 Streptoco |
| 13 | 415.8 | 5.9 | 1443 | 13 | Adr93668 Novel S. |
| 14 | 383.2 | 5.5 | 18274 | 6 | ABA01446 Streptoco |
| 15 | 376.2 | 5.4 | 19966 | 3 | Aaz30355 Nucleotid |
| 16 | 376.2 | 5.4 | 19966 | 3 | Aaz45258 DNA enod |
| 17 | 376 | 5.4 | 28170 | 6 | ABA01447 Streptoco |
| 18 | 373 | 5.3 | 15962 | 6 | ABA01445 Streptoco |
| 19 | 371.6 | 5.3 | 20555 | 6 | ABA01443 Streptoco |
| 20 | 371.4 | 5.3 | 18488 | 6 | ABA01444 Streptoco |

| | | | | | |
|----|-------|-----|-------|----|--------------------|
| 21 | 371.4 | 5.3 | 19738 | 6 | ABA01436 Streptoco |
| 22 | 366 | 5.2 | 25243 | 6 | ABA01442 Streptoco |
| 23 | 364.8 | 5.2 | 22157 | 6 | ABA01448 Streptoco |
| 24 | 347.2 | 5.0 | 14142 | 6 | ABA01435 Streptoco |
| 25 | 346 | 4.9 | 12786 | 6 | ABA01433 Streptoco |
| 26 | 332 | 4.7 | 14460 | 6 | ABA01440 Streptoco |
| 27 | 320.2 | 4.6 | 1386 | 6 | ABN70387 Streptoco |
| 28 | 320.2 | 4.6 | 1386 | 6 | ABN67429 Streptoco |
| 29 | 314.8 | 4.5 | 15347 | 6 | ABA01439 Streptoco |
| 30 | 301.4 | 4.3 | 8056 | 8 | ABZ10246 Haematopo |
| 31 | 298.8 | 4.3 | 8056 | 8 | ABZ10246 Haematopo |
| 32 | 291.6 | 4.2 | 16110 | 6 | ABA01437 Streptoco |
| 33 | 283.8 | 4.1 | 1002 | 10 | ADC50017 N-acetyl |
| 34 | 255 | 3.6 | 447 | 6 | ABN67430 Streptoco |
| 35 | 230.2 | 3.3 | 12494 | 6 | ABA01438 Streptoco |
| 36 | 222.2 | 3.2 | 8056 | 8 | ABZ10100 Haematopo |
| 37 | 219.8 | 3.1 | 8056 | 8 | ABZ10100 Haematopo |
| 38 | 206.6 | 3.0 | 50000 | 6 | ABL55643 AmEPV gen |
| 39 | 198.4 | 2.8 | 15794 | 6 | ABA01434 Streptoco |
| 40 | 187.8 | 2.7 | 5979 | 4 | AAS45313 Chemical |
| 41 | 187.8 | 2.7 | 5979 | 6 | ABK28152 DNA trans |
| 42 | 185.8 | 2.7 | 50000 | 6 | ABL55643 AmEPV gen |
| 43 | 183.6 | 2.6 | 8170 | 6 | ABK28258 DNA trans |
| 44 | 179 | 2.6 | 11691 | 6 | ABL34241 Human imm |
| 45 | 176 | 2.5 | 372 | 8 | ACA50223 Prokaryot |

ALIGNMENTS

RESULT 1

AAZ60930
ID AAZ60930 standard; DNA; 6992 BP.

AC AAZ60930;

DT 30-MAY-2000 (first entry)

DE Nucleotide sequence of a capsular gene cluster of S. suis serotype 1.

XX Capsular gene cluster; serotype 1; polysaccharide biosynthesis;
XX capsular component; antigen; regulation; chain length determination;
XX complement-mediated opsonophagocytosis; serotype-specific detection;
XX antigen; vaccine; Streptococcal disease; CpsIE; CpsIF; CpsIG; CpsIH;
XX CpsII; CpsIJ; CpsLK; glycosyltransferase; Cp polymerase; ss.

OS Streptococcus suis.

| Key | Location/Qualifiers |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| XX | 2..1367 |
| FT | /*tag= a |
| FT | /note= "CpsIE; encodes AAY68970" |
| FT | 1374..1823 |
| FT | /*tag= b |
| FT | /note= "CpsIF; encodes AAY68971" |
| FT | 1823..2317 |
| FT | /*tag= c |
| FT | /note= "CpsIG; encodes AAY68972" |
| FT | 3036..4202 |
| FT | /*tag= d |
| FT | /note= "CpsIH; encodes AAY68973; the nucleotides encoding amino acid 213 are not given" |
| FT | 4195..5163 |
| FT | /*tag= e |
| FT | /note= "CpsII; encodes AAY68974" |
| FT | 5172..6143 |
| FT | /*tag= f |
| FT | /transl_except= (pos: 5634..5636, aa: Xaa) |
| FT | /transl_except= (pos: 5637..5638, aa: Xaa) |
| FT | /note= "CpsIJ; Xaa is an unspecified amino acid; encodes AAV68975; nucleotides 5328-5330 encode an amino acid that is not included in the protein sequence" |
| FT | 6156..6990 |
| FT | CDS |

| | | | |
|----|------|-------------------------------------------------------------------|------|
| Qy | 3721 | TTGCTTGGATAAAAAAGCTAATAGTAATAATTTGTAACTACTATTATTTATTTAAATCTG | 3780 |
| Db | 3721 | TTGCTTGGATAAAAAAGCTAATAGTAATAATTTGTAACTACTATTATTTATTTAAATCTG | 3780 |
| Qy | 3781 | AAATTGCTTTACATGAATTTTGGCTGTTTATAATCTAGAGAAATCAAGTAACGAAGCTA | 3840 |
| Db | 3781 | AAATTGCTTTACATGAATTTTGGCTGTTTATAATCTAGAGAAATCAAGTAACGAAGCTA | 3840 |
| Qy | 3841 | GATTTATTTATTTACAGGAAGTATTGATTAAGTATTAGAAAACAATATTTATTTTGGAT | 3900 |
| Db | 3841 | GATTTATTTATTTACAGGAAGTATTGATTAAGTATTAGAAAACAATATTTATTTTGGAT | 3900 |
| Qy | 3901 | ATGGAATATCCGAATATTCAGTTACGGGAATCTGGCTCGGAAGTCATTCAGGCTATATAT | 3960 |
| Db | 3901 | ATGGAATATCCGAATATTCAGTTACGGGAATCTGGCTCGGAAGTCATTCAGGCTATATAT | 3960 |
| Qy | 3961 | CAATTTTTTATAAATCAGGAATAGTTGGTGTGATTTTACTGATGTTTTCTTTTTTTATG | 4020 |
| Db | 3961 | CAATTTTTTATAAATCAGGAATAGTTGGTGTGATTTTACTGATGTTTTCTTTTTTTATG | 4020 |
| Qy | 4021 | TTATAAAAAAAGTTATGGAGTTAATGGGGAACAGCACTATTTTATTTTACATCATTAG | 4080 |
| Db | 4021 | TTATAAAAAAAGTTATGGAGTTAATGGGGAACAGCACTATTTTATTTTACATCATTAG | 4080 |
| Qy | 4081 | CCATATTTTTCATATATGAACAATAGATCCGATATATATATATATAGTACTATCTTTT | 4140 |
| Db | 4081 | CCATATTTTTCATATATGAACAATAGATCCGATATATATATATATAGTACTATCTTTT | 4140 |
| Qy | 4141 | CTTCAATAGGTATTTTGGAAATATAAATTTTAAAAAGGATATCGAGACAAAAAATCAAT | 4200 |
| Db | 4141 | CTTCAATAGGTATTTTGGAAATATAAATTTTAAAAAGGATATCGAGACAAAAAATCAAT | 4200 |
| Qy | 4201 | GATTTAAATTTTCCAGTTATTTGTACCAATTTATAATGTCCAAGATTATCTTGTAAATGTATT | 4260 |
| Db | 4201 | GATTTAAATTTTCCAGTTATTTGTACCAATTTATAATGTCCAAGATTATCTTGTAAATGTATT | 4260 |
| Qy | 4261 | AACAGTATTTAACCAGAAATATCTAAATTTAGAGGTATTTCTCGTAATATGATGGAAGT | 4320 |
| Db | 4261 | AACAGTATTTAACCAGAAATATCTAAATTTAGAGGTATTTCTCGTAATATGATGGAAGT | 4320 |
| Qy | 4321 | ACTGATGATCTGAGAAAAATTTGCTTAACTATATGAAGACGATGGAAGATTTAAATAT | 4380 |
| Db | 4321 | ACTGATGATCTGAGAAAAATTTGCTTAACTATATGAAGACGATGGAAGATTTAAATAT | 4380 |
| Qy | 4381 | TACAAGAAAAATTAATGGCGGTCTAGCAGATGCTCGAAATTTTCGGACTAGAACATGCAACA | 4440 |
| Db | 4381 | TACAAGAAAAATTAATGGCGGTCTAGCAGATGCTCGAAATTTTCGGACTAGAACATGCAACA | 4440 |
| Qy | 4441 | GGTAAATATATTTGCTTTTGTCCGATTTCTGATGACTATATAGAAAGTTGCAATGTTTCGAGAGA | 4500 |
| Db | 4441 | GGTAAATATATTTGCTTTTGTCCGATTTCTGATGACTATATAGAAAGTTGCAATGTTTCGAGAGA | 4500 |
| Qy | 4501 | ATGCATGATAATAAATCTGAGTATAATGSCCATATAGCAGATAGATTTTGTGTTAGTA | 4560 |
| Db | 4501 | ATGCATGATAATAAATCTGAGTATAATGSCCATATAGCAGATAGATTTTGTGTTAGTA | 4560 |
| Qy | 4561 | GACGAAAAACGGGTATACAAAGAAAAAAGAAATAGTAATTTTTCATGTCTTAAACGAGAGAA | 4620 |
| Db | 4561 | GACGAAAAACGGGTATACAAAGAAAAAAGAAATAGTAATTTTTCATGTCTTAAACGAGAGAA | 4620 |
| Qy | 4621 | GAGACTGTAAAGAAATTTTGTCCAGGATCTAATATAGAAAAATATATAGTATTTGGTCAAGCTT | 4680 |
| Db | 4621 | GAGACTGTAAAGAAATTTTGTCCAGGATCTAATATAGAAAAATATATAGTATTTGGTCAAGCTT | 4680 |
| Qy | 4681 | TATTCACGAGATATTAATAAAGATATAAAATTCCAAATTAATATAGAAATATTTGGTGAG | 4740 |
| Db | 4681 | TATTCACGAGATATTAATAAAGATATAAAATTCCAAATTAATATAGAAATATTTGGTGAG | 4740 |
| Qy | 4741 | GATTTGCTTTTTTAAATTTGGAGGTCTTGAAACAATGTACACGTGTAGTATCTAGTA | 4800 |
| Db | 4741 | GATTTGCTTTTTTAAATTTGGAGGTCTTGAAACAATGTACACGTGTAGTATCTAGTA | 4800 |

| | | | |
|----|------|---------------------------------------------------------------------|------|
| Qy | 4801 | GAATATTTATTAATTTATGTCTTAAAGCAGTTTCCTTATTAATCAGAAATCTCTATA | 4860 |
| Db | 4801 | GAATATTTATTAATTTATGTCTTAAAGCAGTTTCCTTATTAATCAGAAATCTCTATA | 4860 |
| Qy | 4861 | AATAATATTTAGTTAGTCAAGATTTGGAGAAATTTAGCCCTTTTAAGTTTAAAGAGAGATTT | 4920 |
| Db | 4861 | AATAATATTTAGTTAGTCAAGATTTGGAGAAATTTAGCCCTTTTAAGTTTAAAGAGAGATTT | 4920 |
| Qy | 4921 | AGTCATTTATTTGATCGAAAAAGTTTAAAGAGAGAGTTTAAATGTTTAAACAAAAATGTAT | 4980 |
| Db | 4921 | AGTCATTTATTTGATCGAAAAAGTTTAAAGAGAGAGTTTAAATGTTTAAACAAAAATGTAT | 4980 |
| Qy | 4981 | TCAACAGATTTGTTTGGATATAGTCTTTCGCCAATATTTAGAGTCTTATCCGAAAGAAATA | 5040 |
| Db | 4981 | TCAACAGATTTGTTTGGATATAGTCTTTCGCCAATATTTAGAGTCTTATCCGAAAGAAATA | 5040 |
| Qy | 5041 | CGTAGATATCCATTTTAAAGCGAAAAAGATATTTTATCAAGAAAGCAATTTTGTAGTCTTG | 5100 |
| Db | 5041 | CGTAGATATCCATTTTAAAGCGAAAAAGATATTTTATCAAGAAAGCAATTTTGTAGTCTTG | 5100 |
| Qy | 5101 | TATTCATGAAATTTTTCGCCCTAAACTATATATGTAATGTTATATAGAAATTTTCAAAAGCAG | 5160 |
| Db | 5101 | TATTCATGAAATTTTTCGCCCTAAACTATATATGTAATGTTATATAGAAATTTTCAAAAGCAG | 5160 |
| Qy | 5161 | TAGAGCTTAAAAATGATATAAATTTAGTGTATTTCTCCAGTTTATTAATGTAAGATAATTT | 5220 |
| Db | 5161 | TAGAGCTTAAAAATGATATAAATTTAGTGTATTTCTCCAGTTTATTAATGTAAGATAATTT | 5220 |
| Qy | 5221 | TAAAGTATTTGTAAGAAAGCAATTTTAAATCAAAATTTATAAAAATATAGAAATATTTATGA | 5280 |
| Db | 5221 | TAAAGTATTTGTAAGAAAGCAATTTTAAATCAAAATTTATAAAAATATAGAAATATTTATGA | 5280 |
| Qy | 5281 | TAGATGATGGCTCTGTAGATGATTTCTGTAAATAATATGCAAGAAATATGCAAGAAAGATA | 5340 |
| Db | 5281 | TAGATGATGGCTCTGTAGATGATTTCTGTAAATAATATGCAAGAAATATGCAAGAAAGATA | 5340 |
| Qy | 5341 | AAAGCTTAAAAATTTTTCACCTAAATCAGTGGAGTATCAAAATGCTAGAAATCATGGAA | 5400 |
| Db | 5341 | AAAGCTTAAAAATTTTTCACCTAAATCAGTGGAGTATCAAAATGCTAGAAATCATGGAA | 5400 |
| Qy | 5401 | TAAAGCGGAGTACAGCTGAATATTTATGTTGTTGCTGACTCTGATGATGTTTGTGATAGTA | 5460 |
| Db | 5401 | TAAAGCGGAGTACAGCTGAATATTTATGTTGTTGCTGACTCTGATGATGTTTGTGATAGTA | 5460 |
| Qy | 5461 | GATTAGTAGAAAAATTTATTTTAAATATTTATAAAAAAGTAGAAGTATTTCTGGTTGTT | 5520 |
| Db | 5461 | GATTAGTAGAAAAATTTATTTTAAATATTTATAAAAAAGTAGAAGTATTTCTGGTTGTT | 5520 |
| Qy | 5521 | TGTACGCTACTTTTTCAGAAAAATATAAATAATTTTGAAGTGAATATCCCAATATTTGATTT | 5580 |
| Db | 5521 | TGTACGCTACTTTTTCAGAAAAATATAAATAATTTTGAAGTGAATATCCCAATATTTGATTT | 5580 |
| Qy | 5581 | TTCAAGCAATTAATAACCGTCAGGACATGCGAGAAAAAATTTTATGAAATTTGTATATAA | 5640 |
| Db | 5581 | TTCAAGCAATTAATAACCGTCAGGACATGCGAGAAAAAATTTTATGAAATTTGTATATAA | 5640 |
| Qy | 5641 | ATAATATTTTTTCTACTCTCTGTTTAAACTATAATAAGAAAGATACATAACAGATCTTT | 5700 |
| Db | 5641 | ATAATATTTTTTCTACTCTCTGTTTAAACTATAATAAGAAAGATACATAACAGATCTTT | 5700 |
| Qy | 5701 | TTCAAGAGATCAATGTTAGGAGAGATTTTATCTTTTATCTGCAATTTTAAAGAAATA | 5760 |
| Db | 5701 | TTCAAGAGATCAATGTTAGGAGAGATTTTATCTTTTATCTGCAATTTTAAAGAAATA | 5760 |
| Qy | 5761 | TAGATAGAGTTAGTATTTTACCTGAACTCATCTTTTATTTTATAGGAGAGGTATCTAAGTA | 5820 |
| Db | 5761 | TAGATAGAGTTAGTATTTTACCTGAACTCATCTTTTATTTTATAGGAGAGGTATCTAAGTA | 5820 |
| Qy | 5821 | CAGTAAATCTTTTAAAGAGAGGTGTGTTTTGCAATTTGGAAAAATTTTGCACAAAAACAAGTGA | 5880 |
| Db | 5821 | CAGTAAATCTTTTAAAGAGAGGTGTGTTTTGCAATTTGGAAAAATTTTGCACAAAAACAAGTGA | 5880 |
| Qy | 5881 | TAGTATTTGTTTAAAGCAAAATATATGTTGAGGATTTTTCAGCGTATCAATTTGTTTAAAGATACTA | 5940 |

[illegible]

Db 6961 ATCATTTGGTTGAGCAAAACAGCAAAAGAGCTT 6992

RESULT 2
AAZ60929
ID AAZ60929 standard; DNA; 26281 BP.
XX AC AAZ60929;
XX
DT 30-MAY-2000 (first entry)
XX
DE Nucleotide sequence of a capsular gene cluster of *S. suis* serotype 2.
XX
KW Capsular gene cluster; serotype 2; polysaccharide biosynthesis;
KW capsular component; antigen; regulation; chain length determination;
KW complement-mediated opsonophagocytosis; serotype-specific detection;
KW antigen; vaccine; Streptococcal disease; ORF 22; ORF 2Y; ORF 2Z; Cps2A;
KW Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J; Cps2K;
KW Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T; ss.
XX
XX Streptococcus suis.
XX
FH Location/Qualifiers
FT 2..721
FT /tag= a
FT /note= "ORF 2Z; encodes AAY68950"
FT complement (822..2079)
FT /tag= b
FT /note= "ORF 2Y; encodes AAY68951"
FT 2202..2936
FT /tag= c
FT /transl_except= (pos: 2916..2918, aa: Xaa)
FT /note= "ORF 2X; Xaa is an unspecified amino acid; encodes
FT AAY68952"
FT 3041..4486
FT /tag= d
FT /note= "Cps2A; encodes AAY68953"
FT 4504..5193
FT /tag= e
FT /note= "Cps2B; encodes AAY68954"
FT 5203..5880
FT /tag= f
FT /note= "Cps2C; encodes AAY68955"
FT 5919..6650
FT /tag= g
FT /note= "Cps2D; encodes AAY68956"
FT 6675..8054
FT /tag= h
FT /note= "Cps2E; encodes AAY68957"
FT 8089..9258
FT /tag= i
FT /note= "Cps2F; encodes AAY68958"
FT 9262..10419
FT /tag= j
FT /transl_except= (pos: 10057..10059, aa: Xaa)
FT /note= "Cps2G; Xaa is an unspecified amino acid; encodes
FT AAY68959"
FT 10808..12176
FT /tag= k
FT /transl_except= (pos: 11963..11965, aa: Xaa)
FT /note= "Cps2H; Xaa is an unspecified amino acid; encodes
FT AAY68960"
FT 12213..13445
FT /tag= l
FT /note= "Cps2I; encodes AAY68961"
FT 13583..14581
FT /tag= m
FT /note= "Cps2J; encodes AAY68962"
FT 14574..15578
FT /tag= n
FT /note= "Cps2K; encodes AAY68963"
FT 18401..18904
FT /tag= o
FT


```
FT /transl_except= (pos: 18755..18957, aa: Ile)
FT /transl_except= (pos: 18770..18772, aa: Leu)
FT /transl_except= (pos: 18777..18778, aa: Ile)
FT /transl_except= (pos: 18836..18838, aa: Pro)
FT /transl_except= (pos: 18890..18892, aa: Thr)
FT /transl_except= (pos: 18896..18898, aa: Pro)
FT /transl_except= (pos: 18935..18937, aa: Pro)
FT /transl_except= (pos: 18953..18955, aa: Leu)
FT /transl_except= (pos: 18968..18970, aa: Xaa)
FT /transl_except= (pos: 18971..18973, aa: Ser)
FT /transl_except= (pos: 18977..18979, aa: Gln)
FT /transl_except= (pos: 18980..18982, aa: Asn)
FT /transl_except= (pos: 18989..18991, aa: Asn)
FT /note= "Cps20; Xaa is an unspecified amino acid; encodes
FT AAY68964"
FT CDS 20327..21343
FT /*tag= p
FT /note= "Cps2P; encodes AAY68965"
FT CDS 21355..21867
FT /*tag= q
FT /note= "Cps2Q; encodes AAY68966"
FT CDS 21933..22487
FT /*tag= x
FT /note= "Cps2R; encodes AAY68967"
FT CDS 22501..23127
FT /*tag= s
FT /note= "Cps2S; encodes AAY68968"
FT CDS 23136..24368
FT /*tag= t
FT /note= "Cps2T; encodes AAY68969"
FT XX
FT WO200005378-A2.
FT XX
FT 03-FEB-2000.
FT XX
FT 19-JUL-1999; 99WO-NL000460.
FT XX
FT 22-JUL-1998; 98EP-00202465.
FT PR 22-JUL-1998; 98EP-00202467.
FT XX
FT (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
FT PA
FT Smith HE;
FT DI
FT WPI; 2000-195104/17.
FT DR
FT P-PSDB; AAY68950, AAY68951, AAY68952, AAY68953, AAY68954, AAY68955,
FT DR AAY68956, AAY68957, AAY68958, AAY68959, AAY68960, AAY68961, AAY68962,
FT DR AAY68963, AAY68964, AAY68965, AAY68966, AAY68967, AAY68968, AAY68969.
FT XX
FT New nucleic acid containing the capsular gene cluster of Streptococcus
FT PT suis, used for serotype-specific detection and to generate antigens or
FT PT mutants for vaccination.
FT XX
FT Claim 4; Fig 3; 144pp; English.
FT PS
FT XX
FT The present sequence represents the capsular gene cluster of
FT CC Streptococcus suis serotype 2. The genes in this cluster are involved in
FT CC polysaccharide biosynthesis of capsular components and antigens. The
FT CC proteins are involved in regulation (CpsA), chain length determination
FT CC (CpsB, CpsC), export (CpsC), and biosynthesis (CpsE, CpsF, CpsG, CpsH,
FT CC CpsJ, CpsK). The capsule confers bacterium resistance to complement-
FT CC mediated opsonophagocytosis. The gene cluster is used as a source of
FT CC probes and primers for serotype-specific detection of S. suis and is also
FT CC useful for recombinant production of the proteins. The proteins are then
FT CC useful for producing antigens that can be used in vaccines, for
FT CC controlling or eradicating a Streptococcal disease, in humans or animals,
FT CC e.g. against S. suis in pigs
FT XX
FT SQ Sequence 26281 BP; 8252 A; 4158 C; 5488 G; 8383 T; 0 U; 0 Other;
FT Query Match 17.1%; Score 1195.8; DB 3; Length 26281;
FT Best Local Similarity 78.9%; Pred. No. 1.1e-151;
FT Matches 1437; Conservative 0; Mismatches 382; Indels 2; Gaps 1;
```

```
Qy 5172 ATGGAATAAATTAGTGTATTGTTCCAGTTTATAAATGTTAGATAAATAATTTAAGTAGTTGT 5231
Db 13583 ATGGAAAAATCAGCATTTATTTGACCTATTTTAAATACGGAAGTACTTTAAGAGAGTGT 13642
Qy 5232 ATAGAAGCATTTAATAATCAAAATTAATAAATAATAGAAATATTTATGATAGATGATGGC 5291
Db 13643 TTAGATAGCATTTATTTCCCAATCGTATACATACTAGATTCTTTTGATAGATGACGGT 13702
Qy 5292 TCTGTAGATGATTCTGCTAAAAATATGCAAGGAATATGCGAAAAAAGATAAAGAGTAAAA 5351
Db 13703 TCTTCAGATTCAACAACGGATATATGTTTGGATACGCGACGACAGATGCTAGATAAAA 13762
Qy 5352 ATTTTTCATTAATCATAGTGGAGTATCAAAATGCTAGAAATCATGGAATAAAGCGGAGT 5411
Db 13763 CTTTTCGGTTTACCAATCGTGTGTTTCAACGCAAGGAATTTACGGTATCAAAAAATGAC 13822
Qy 5412 ACAGCTGAATATATTTGTTGTTGACTCTGATGATGTTGTTGATAGTACTAGATTAGTAGAA 5471
Db 13823 ACAGCAAAATTAATATGTTTGTAGATTCTGTAGATATTGTTGACGCAACATTTGTTGAG 13882
Qy 5472 AAATTATATTTTAATATATAAAGTAGAAGTGAATTTATCTGTTGTTGTTGATCGTACT 5531
Db 13883 TCCTTATACACCTGTTTAAAGAGAAATGATAGTGAATTTGTCGGAGGGTTACTTGTCTACT 13942
Qy 5532 TTTTCAGAAAAATATAAATAATTTTGAAGTGAATAATCCAAATAATTTGATTTTGAAGCAATT 5591
Db 13943 TTTGATGGAATTTATCAAGAAATCTGAGCTGCAAAAGTGTCAAAATTTGATTTTGGAAAGATA 14002
Qy 5592 AATACCGTCAGGACATGGGAGAAAAAATTTTATGAAATTTGTTATATATAAATAATATTTT 5651
Db 14003 AAAGAGTGGAGACTTAGGAAATGAAATTTTCCCAATCATTTATATGACGGGTATCTTT 14062
Qy 5652 TCTACTCCTGTTTGTAAACTATATAAGAAAAAGATACATACAGATCTTTTTCGAAGAAAT 5711
Db 14063 AATAGCCCTGTTGCAAACTTTATAGATATATATATAAACCAAGTTTTCACACTGAA 14122
Qy 5712 CAATGTTAGGAGAAGATTTACTTTTAACTGCAATTTTAAAGAAATATAGATAGATAGT 5771
Db 14123 CAGTGGTTAGGAGAGGACTTATTATTAACTTAAATTTTAAAGAAATATAAAAAAGTC 14182
Qy 5772 AGTTATTTGACTGAACATCTTTATTTTATAGAGAGAGGTATACTAGTACAGTAAATTTCT 5831
Db 14183 CGCTATGTTAACAGAAATCTTTATTTTGCAGAAAGAGTTTACAAAGTACTACAAATACG 14242
Qy 5832 TTTTAAAGAAAGTGTGTTTTCGCAATTTGCAAAATTTGCAAAAAACAAGTATAGTATTTCTTT 5891
Db 14243 TTTAATATGATGCTTTTATTCAATTAGAAAAATTTAGAGAAAAAATCTTTTGATTTGTTT 14302
Qy 5892 AAGCAAAATATCGTGAGGATTTTGACGTATCAATTTGTTTAAAGATATATACGTTGGCAA 5951
Db 14303 GTTAAAAATATTGGTGGCAATATGAATTTTCTGTTTAAAGAGACGCTACAGTGGCAT 14362
Qy 5952 GTATTTTATATAGCTTACTTAATGTTTAAATACGGAACACAGTCTATTTTTCACAAATTT 6011
Db 14363 ATTATTTATATAGCTTATTAAATGTTTCAAAATAGGAGATGAATCGCTTCCAAAGAAATG 14422
Qy 6012 TTAATTTTGAAGAACTTTTATAAAAAATATTTTAACTTTGTTTAAAGTATCTAACAAA 6071
Db 14423 CATATATTT--AGTATTTATACAATAGGCAATCTTTTAGATACTCTAAGTATTTAAACGAA 14480
Qy 6072 AATTCCTTTGCTAAAAATTTTGTATAAGAAATTTGTTTCGAACAAAGTTTAAAAAATA 6131
Db 14481 CGTCCCTCTGTTTAAAAAGAAATATGAAATTTAAATTTGTTGCTAATAATTTTGTAAAAATTT 14540
Qy 6132 TTATGGTTAATAAGGAGATATCATGATACTATTAGTAAATTTCTATAATTTGTACCT 6191
Db 14541 TTTTAAATCTTTAAATAGGGAAGAAAAAATTAATGATTAACATTTCTATCATCGTCCCA 14600
Qy 6192 ATATATAATGTAGAAAAATATTTTATCTAAATGTATAGATAGCAATTTGTAATTCAGACCTAC 6251
Db 14601 ATTTACAATGTTGAACAATATCTATCCAAGTGTATAAATAGCAATTTGTAATTCAGACCTAC 14660
```



```
QY 6252 AAACATATAGAGATTCCTCGTGAATGACGGTAGTACGGATAATTCGGAAGAAATTTGT 6311
Db 14661 AAACATATAGAGATTCCTCGTGAATGACGGTAGTACGGATAATTCGGAAGAAATTTGT 14720
QY 6312 TTAGCATATCGAAGAAAGATAGTCGCAATTCGTTATTTTAAAAAGAGAAACGGCGGGCTA 6371
Db 14721 TTAGCATATCGAAGAAAGATAGTCGCAATTCGTTATTTTAAAAAGAGAAACGGCGGGCTA 14780
QY 6372 TCAGATGCCGTAATTTATGGCATAAGTCGCGCCCAAGGGTGACTACTTAGCTTTTATAGAC 6431
Db 14781 TCAGATGCCGTAATTTATGGCATAAGTCGCGCCCAAGGGTGACTACTTAGCTTTTATAGAC 14840
QY 6432 TCAGATGATTTTATTCATTCGAGTTTCATCAACGTTTACAGGAAGCAATTCAGAGAGAG 6491
Db 14841 TCAGATGATTTTATTCATTCGAGTTTCATCAACGTTTACAGGAAGCAATTCAGAGAGAG 14900
QY 6492 AATGCCCTTGTGGCAGTTGCTGTTATCATAGGTAGGTAGTCTCGGGGCATTTCTTAACA 6551
Db 14901 AATGCCCTTGTGGCAGTTGCTGTTATCATAGGTAGGTAGTCTCGGGGCATTTCTTAACA 14960
QY 6552 GCAGAGCGCTTCTTACAAATCAGGCTGTTCTGAGCGGCAGGAATGTTGTAAAAAGCTG 6611
Db 14961 GCAGAGCGCTTCTTACAAATCAGGCTGTTCTGAGCGGCAGGAATGTTGTAAAAAGCTG 15020
QY 6612 CTAGAGCGGATGTCATCGCTTTGTGTCGCTGTATAAATCTATAAAAAAGAACTA 6671
Db 15021 CTAGAGCGGATGTCATCGCTTTGTGTCGCTGTATAAATCTATAAAAAAGAACTA 15080
QY 6672 TTTGAGATTTTCCGATTTGAAAGGTAAGATTCATGAGATGATATCTTACATTCGCTATCGC 6731
Db 15081 TTTGAGATTTTCCGATTTGAAAGGTAAGATTCATGAGATGATATCTTACATTCGCTATCGC 15140
QY 6732 TTGCTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGGAGTGCTGTACTATTATTGTTGAC 6791
Db 15141 TTGCTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGGAGTGCTGTACTATTATTGTTGAC 15200
QY 6792 CGAGAAATAGTATCACAACTTCTAGCATGACTGACCAATCGCTTCCATTCGCTACTGAA 6851
Db 15201 CGAGAAATAGTATCACAACTTCTAGCATGACTGACCAATCGCTTCCATTCGCTACTGAA 15260
QY 6852 TTTCAAATGACGATGACTTCTATGAAAGTAGAGAGATAAGAGCTCTTACTAGAG 6911
Db 15261 TTTCAAATGACGATGACTTCTATGAAAGTAGAGAGATAAGAGCTCTTACTAGAG 15320
QY 6912 TGTATCGTTCATTTTATGACCTTCTGCTTTTGTGTTTATGAGCAATATAATCATTTGGTTG 6971
Db 15321 TGTATCGTTCATTTTATGACCTTCTGCTTTTGTGTTTATGAGCAATATAATCATTTGGTTG 15380
QY 6972 AGCAAAACAGCAAAAGAGCTT 6992
Db 15381 AGCAAAACAGCAAAAGAGCTT 15401
```

RESULT 3

```
ABA01441
ID ABA01441 standard; DNA; 17468 BP.
XX AC ABA01441;
XX AC ABA01441;
XX DT 21-FEB-2002 (first entry)
XX DE Streptococcus thermophilus eps3 operon #2.
XX KW Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
XX KW ds.
XX OS Streptococcus thermophilus.
XX PN W0200179500-A2.
XX PD 25-OCT-2001.
XX PF 18-APR-2001; 2001WO-FR001199.
```

XX

18-APR-2000; 2000FR-00004972.

XX

(INRG) INRA INST NAT RECH AGRONOMIQUE.
(DANO-) CIE DANONE SA GERVAIS.
(RHOD) RHODIA CHIM.

XX

Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;

PI

WPI; 2002-017616/02.

XX

New nucleic acid fragments containing exopolysaccharide operon, useful
e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.

XX

Claim 9; Page 75-80; 144pp; French.

XX

The present sequence is an eps operon from *Streptococcus thermophilus*.
Proteins encoded by the eps operon function in exopolysaccharide (EPS)
synthesis. The operon is useful for producing chimeric eps operons, for
optimising production of EPS in lactic acid bacteria. EPS impart texture,
mouth feel and rheological properties to fermented food products (e.g.
yoghurt). They function as thickeners, to provide free-flowing and creamy
texture, and may also have biological activities beneficial to health

XX

SQ Sequence 17468 BP; 5871 A; 2455 C; 3307 G; 5835 T; 0 U; 0 Other;

Query Match

13.7%; Score 955.6; DB 6; Length 17468;

Best Local Similarity 65.8%; Pred. No. 1.9e-119;

Matches 1484; Conservative 0; Mismatches 729; Indels 41; Gaps 5;

QY

1 ATCCCAAAAGCAATTTGGCAATTTTGTATATGATAGTAGGTTGCAATTTCTGCAATCTTAA 60

Db

6294 ACCGGCAATTTAGACTGGCGCTTTTAGAATTTGATTTGGGGTAGTAGTGGCTGCTCCATTA 6353

QY

61 CAAGTCATATACCAAAATGCTGATTTAATCTGTTCTGGAAATTTTATCATATATGATGTTTC 120

Db

6354 CCACCCACTTGGCCACCCTCAAGATGAATAGTAGGGACGCTATTTGATTTATTTGCCCTGC 6413

QY

121 ATATTTTGTCAATTTTATATCTCGTAGCCAGTTGAATTTTGTAGTATAGAGTATATCTGA 180

Db

6414 ATTTTGTGGCTTTTATTTCTCAGTATGTCATCAGAAATTTTGAGTAGAGGCTACTTAA 6473

QY

181 TAGAGTTTGAATAAACATTTAACTATATAGTATTAATTTTGCAATTTTCTTACGGCAGTAT 240

Db

6474 TAGAGTTTGAACGGTCGCTAGGTATGTTTGTATTTTCGATTTCTCTCATTTTCAATTT 6533

QY

241 CATTTTGTGGAGAAATTTTCGACTTTCAAGACGTGGTGCCGTGATTTTCACATTTAA 300

Db

6534 CTTTTTTATTTGGGAGATAGTTTATCATTTACGACGCTGGGTGGCATTTATTTACCATTC 6593

QY

301 TAAACTTTCGTTTTCGTATACCTATTTAACTGTAATTTTAAAGCAGTTTAAAGGATAGCTTTC 360

Db

6594 TTACCTTTATCATCGTTTATTTATTAATCACTGCATGTTCTGTAGATACAAATACCTTCTCT 6653

QY

361 TATTTTCGCAATCTATCAAAAAAGAGCAATTTCTAATTTACAAACGCTGAAACGATGGAAA 420

Db

6654 TTATGACTGCTGAACACAGAGAATACTTTAGTCAATTTACCACACTACTCAGAGATTCGAT 6713

QY

421 ATATGCAAGTTTATTTGAATCACAATAACAAATTTCAAAAAAATCTTGTTCGATTTGGTAG 480

Db

6714 TAATGGAGGGGCTCTTTGAATCGGATCAACTTCTCCATAATATTTTGGCGGGTATTTAG 6773

QY

481 TTTTAGGTACAGAAATAGATAAATTAATTTTATCATTTACCGCTCTATTTTCTGTGGAAG 540

Db

6774 TCATTTGG---AGACGGTATGTTGCCCTTCCAGAGGGAGTTTCCAGTAGTTCTCTTTTGAAT 6830

QY

541 AAGCTATAGAGTTTTCACAGGGAAGTGGTTCGACCGCTCTTTTATATAATCTTACCAGTG 600

Db

6831 ATGCAATAGAGTTTTCGACCCCATGAGTTTTCGACCATCTGTTTATCACTTACCTACCGAGTG 6890

QY

601 AGTTTATAGCGTAAAGCAATTCGTTTTCAGATTTTGTAGTGTGTTAGTATTCATGTAACG 660

Db

6891 AACATTACGATCTCAAAACATCTTGTTCGATTTTGAAGTTCATGGGTATTCATGATGAGTG 6950


```
Qy 661 TTGATATTAATTCATTTCGGTTTACTCGTTGAAACACAAAATCCAACTGCTAGGTG 720
Db 6951 TAGATATTAACACATCGATTCGTTTAAAGGCTTTAAACAAAATAAAAAATCAACAGGTTGGAG 7010
Qy 721 ACCATAGCATTTGTAACCTTTTCCACAAATTTTATAAGCCCTAGTCATATCATGATGAAC 780
Db 7011 ACCATAGCATCGTGACCTTTAACTCCAAATTAATACAAACATAGCCATATCTTCTAAAGC 7070
Qy 781 GACTTTTGGATATACCTCGGAGCGGTAGTCGGGTTAAATTAATTTGGTGTATGTTCTATTT 840
Db 7071 GCATGTTGGATATCTTTGGGCGCTTGTATGCTCTCTAAATTTGGTGTCTGCTGGGATTTG 7130
Qy 841 TGTGTAGTTCCAAATTAATTCGTAGATGCTGGACCGGCTATTTTGTCTCAGAAACGAGTTG 900
Db 7131 TCTTAGCCCCCATCAITTCGGAAGATGGAGACCGGCCATTTTGTTCAGAAACGAGTAG 7190
Qy 901 GACAGAAATGAGCGCATATTTACATTTCTACAAGTTTCGATCGATGATGTTGATGCTCAGG 960
Db 7191 GGAATAACGGACGTATCTTTAAATTTTAAATTCGGTTCTATGTACACTGATGCGGAGG 7250
Qy 961 AGCCCAAAAAAGACTGCTCAGCCCAAAACCAAGATGCAAGGGTGGGTATGTTTAAAAATGG 1020
Db 7251 A---AAGAAAGAAATGAAAGCACAAAATCAGATGCA-----CGGTGGAATGTTCAAGAT 7301
Qy 1021 GAAAAACGATCCTAGAAATTAATCAATTTCAAGTTTCGATCGATGATGTTGATGCTCAGG 1080
Db 7302 GAATAACGATCCAGTATCACTCTATCGGTCGTTTATTCG-AAAAACGAGTCTGGAGC 7360
Qy 1081 AGTTACACAGTTTAAATGTTTAAATGGCGGATATGAGTCTAGTTAGTTGACAGCTCCAC 1140
Db 7361 AACTGCTCAGTTTAAATGTTTAAATGTTGGTTCGGGATATGAGTCTTGTGGGACTCGCCAC 7420
Qy 1141 CTAAGTTGATGAATTTGAAAAATATATCTCTGTCGTCAAAAAGACAGATGAGTTTAAAC 1200
Db 7421 CAACGTGCGATGAATTTGAAAAATATACACCAAGTCAAAACGTCGCTCAGCTTCAAGC 7480
Qy 1201 CAGGGATACAGGCTCTGGCAGGTTAGTGTCTAGTATATATACAGACTTCGACGACG 1260
Db 7481 CAGGAATACAGGCTATGCGCAGCAAGTGTGCTGATTAATATACCAACTTTTGATGAGG 7540
Qy 1261 TAGTTCGGTTGGAATAGCATATGATATTCGACTATCTGCTCAGATATTAATAAATTT 1320
Db 7541 TGGTAAGACTCGATGTTGAATATATCGATATCTGCTGATTTGGTTCGATATCAAGATAT 7600
Qy 1321 TATTAAGACAGTGAAGTTGTAATGTTGAGAGGGAAGTAAAGT----- 1365
Db 7601 TGTTGAAGACTATTTGTCGTGTTTAAAGAAAGGGAAGTAGGTAGAGCCCCCTCTTCA 7660
Qy 1366 -----AAAAGTATATGAAGTTTGTGTTGAGAGGGAAGTAAAGT----- 1365
Db 7661 TTCTTGTATTGGGAGAAAAAATGAAATTTGTTTAGTGGGTCACTCGGAGGCAATTTA 7720
Qy 1413 ACTCAGTTGATTTGTTTAAACCGTTTGGAGGAAGAACGCTTTTGGGTAAACATTT 1472
Db 7721 GCATTTTAAATATGTTTAAACCAATCTGGAGAGATGAAGAACGTTTGGTGAACCTTT 7780
Qy 1473 GATAAAGAGGATCAAGAAGCTTTTGAAGAAATGAAAAAATGATTCATGTTACTTTCCA 1532
Db 7781 GATAAAGAGGATCAAGAAGCATCTTAAAGATGAAAAAATGATTTCTGTCACCTTTCT 7840
Qy 1533 ACAATCGCATCTCATTAATTTAGTGAATAATCTTTCTAGCTTTCAAAATTTTACGT 1592
Db 7841 ACAATCGAACTTTAAAAAATTTAAATTAATAAATCTTTTCTAGCTATTAAGTATTAATA 7900
Qy 1593 GATGAGAAACAGATGTTATTTATTCATCTGTCGCGCGTGTGTCCTCTCTTTTAC 1652
Db 7901 AAAGAAAAACCTGATGTCATTTATTTCTGAGACTGCGTAGCCGTTCCATCTTTTAC 7960
Qy 1653 ATCGGAAACATTTTGGAGCAAGACGATTTATATTGAAGTATTTGATCGAGTTTAAATA 1712
Db 7961 ATTCGTAACATTTTGGAGCTAAGACGCTTTATATTGAAGTCTTTGTATAGATTAATAA 8020
```

```
Qy 1713 TCTACATTAACCTGAAAACTAGTTTATCCGTAACAGATATTTTATTGTTCTAGTGGAA 1772
Db 8021 CCAACAGTTACAGAAAAATTAGTGTACCTGTGCGGATAGTTCATTTCAATGGAA 8080
Qy 1773 GAAATGAAGAAGGTATATCCTAAATCTATTAATCTATTAACCTGCGGAGTATTTTTTAATGATTTT 1832
Db 8081 GAGATGAAAAAAGCTTTATCCAAAAGCTATTAATTTAGGAGTATCTTTTAAATGATTTT 8140
Qy 1833 TAAACAGTAGAATCATGAAACAAACAGTTTAATCGATGATAAAGAGATTTGATTTATGA 1892
Db 8141 TAAACAGTAGGAACACACGAAACAAACCTTTTAAATCGCTTAATTTAGGAAGTGGATCATTTAG 8200
Qy 1893 AAAAAATGAAAGTATAAACCAGCAAAATATTTTATCAAAACAGGATATCTCTGACTATATTC 1952
Db 8201 TTGAAAACAGGAGTGATAAAGGAAGAGGTCTTTTATCAACAGGATATTTCTATCTACGAAC 8260
Qy 1953 CAGAATATTGCAAGTATAAAAAATTTCTCAGTTTACAAAGAAATGGAACAAATATATTAACA 2012
Db 8261 CTAAATTTTGTGAGTGTCTCGTTTGTATAGTTTGTGATCAAAATGAGTAAATTTATGCAAA 8320
Qy 2013 AATCAGAAGTAGTTATTTGCGACGAGGCCCGCTACTTTTATGAAATTCATTTATCCAAAG 2072
Db 8321 AAGCTGATATTTATATCACTCAGGAGGTCAGCGACTTTTATGAGTGCCATTAACAATG 8380
Qy 2073 GAAAAAACAATTTATTTGTTTCTAGACAAAAGAAAGTATGTTGAAATGATCATC 2132
Db 8381 GCAAAAGCCAAATTTGTTGTCCTAGACAAAAGAAAGTATGTTGAAATGATCATC 8440
Qy 2133 AAGTAGAGTTTGTAGAAGAAATTTTACAAGATAATAATTTTATTTATAGAAAAATATAG 2192
Db 8441 AAGTAGATTTTGTAGAAGAAATTTTACAAGATAATAATTTTATTTATAGAAAAATATAG 8500
Qy 2193 ATGATTTGTTTGAATAAATTTATTTGAAGTTTCTAA 2226
Db 8501 TAGATAAATTTGGGGAGACTATTTGAAAAATTTACAA 8534

RESULT 4
ID ADS13200 standard; DNA; 16032 BP.
XX
AC ADS13200;
XX
XX
DT 16-DEC-2004 (first entry)
XX
DE S. thermophilus CNCM I-2980 polysaccharide synthesis-related operon DNA.
KW lactic acid bacterium; food; meat; cereal; dairy;
KW polysaccharide synthesis; operon; ds.
XX
OS Streptococcus thermophilus.
XX
FH Key Location/Qualifiers
FT CDS 342..1802
FT /tag= a
FT /product= "Transcriptional regulator eps13A protein"
FT /note= "The corresponding protein sequence is not shown
FT shown within the specification"
FT CDS 1803..2534
FT /tag= b
FT /product= "Polysaccharide polymerisation/export-related
FT eps13B protein"
FT /note= "The corresponding protein sequence is not shown
FT shown within the specification. The CDS has a prokaryotic
FT Val start."
FT CDS 2543..3235
FT /tag= c
FT /product= "Polysaccharide polymerisation/export-related
FT eps13C protein"
FT /note= "The corresponding protein sequence is not shown
FT shown within the specification"
FT CDS 3245..3985
FT /tag= d
```



```
FT FT /product= "Polysaccharide polymerisation/export-related
FT FT eps13D protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 4042..5409
FT FT /tag= e
FT FT /product= "Undecaprenyl-phosphate glycosyltransferase
FT FT eps13E protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification. The CDS has a prokaryotic
FT FT Val start."
FT CDS 5611..6195
FT FT /tag= f
FT FT /product= "Undecaprenyl-phosphate glycosyltransferase
FT FT eps13F protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification. The CDS has a prokaryotic
FT FT Val start."
FT CDS 6251..6634
FT FT /tag= g
FT FT /product= "Undecaprenyl-phosphate glycosyltransferase
FT FT eps13G protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 6643..7092
FT FT /tag= h
FT FT /product= "Beta-1,4-galactosyltransferase eps13H protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 7092..7607
FT FT /tag= i
FT FT /product= "Beta-1,4-galactosyltransferase eps13I protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 7597..8493
FT FT /tag= j
FT FT /product= "Rhamnosyltransferase eps13J protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification. The CDS has a prokaryotic
FT FT Val start."
FT CDS 8763..9797
FT FT /tag= k
FT FT /product= "Glycosyltransferase eps13K protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 9827..10969
FT FT /tag= l
FT FT /product= "Repetitive unit polymerase eps13L protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 10984..11793
FT FT /tag= m
FT FT /product= "Repetitive unit polymerase eps13M protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 11844..12578
FT FT /tag= n
FT FT /partial
FT FT /product= "Glycosyltransferase eps13N protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification. No start codon."
FT CDS 12633..13016
FT FT /tag= o
FT FT /product= "Glycosyltransferase eps13O protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS 13049..14482
FT FT /tag= p
FT FT /product= "Transmembrane transporter eps13P protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification. The CDS has a prokaryotic
FT FT Val start."
FT CDS complement(14614..15870)
```

```
FT FT /tag= q
FT FT /product= "Transposase IS1193 protein"
FT FT /note= "The corresponding protein sequence is not shown
FT FT shown within the specification"
FT CDS FR2852604-Al.
FT FT 24-SEP-2004.
FT FT 17-MAR-2003; 2003FR-00003242.
FT FT 17-MAR-2003; 2003FR-00003242.
FT FT (RHOD ) RHODIA CHIM SAS.
FT FT Horvath P, Manoury E, Huppert S, Fremaux C;
FT FT WPI; 2004-671092/66.
FT FT New strain of lactic acid bacteria, useful for preparation of foods and
FT FT pharmaceuticals, especially fermented dairy products, contains genes
FT FT involved in polysaccharide biosynthesis.
FT FT Claim 5; SEQ ID NO 1; 35pp; French.
FT FT The invention relates to a novel strain of a lactic acid bacterium that
FT FT contains at least one of 8 specific nucleic acid sequences all of which
FT FT are defined in the specification. The bacterium of the invention may be
FT FT useful during food preparation, particularly that of beverages and meat,
FT FT cereal, or dairy products e.g. fermented milk, yoghurt, 'matured' cream,
FT FT dessert cream, cheese, soft cheese, cheese spreads, cottage cheese, milk-
FT FT based drinks, dairy product tentates and baby milk, where the products
FT FT are derived from an animal and/or plant. The current sequence is that of
FT FT the Streptococcus thermophilus CNCM I-2980 polysaccharide synthesis-
FT FT related operon DNA of the invention.
FT SQ Sequence 16032 BP; 5191 A; 2323 C; 3188 G; 5330 T; 0 U; 0 Other;
FT Query Match 12.7%; Score 889; DB 13; Length 16032;
FT Best Local Similarity 66.8%; Pred. No. 1.6e-110;
FT Matches 1344; Conservative 0; Mismatches 655; Indels 12; Gaps 5;
FT QY 317 ATACCTTTTAAACGTAATTAATTAAAGCAGTTTAAAGTAGCTTCTTATTTTCGACAATCTA 376
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5592 ATACTTTATCACTGCTGCTGCGTAGATACAAATATCTATTTCTTATGACTGCTGAACA 5651
FT QY 377 TCAAAAAGACGATTTCTAATATACACGGCTGAACGATGGGAAATATGCAAGTTTTATT 436
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5652 ACAGAAGAATACATTAGTCATTACCACCTGAGAGATTGCGATTAAATGGAGGGGCTCTT 5711
FT QY 437 TGAATCATATAACAAATTTCAAAAATCTTTGTCATTGTTAGTGTGTTAGGTACAGAAAT 496
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5712 TGAGTCAGATCAACTTCTCTATCAAAATATTGGCAGGTATAGTAGTCATCGGAGATGGTGA 5771
FT QY 497 AGATAAAATTAATTTATCATTACCGCTCTATTATTCTGTGAAGAAGCTATAGAGTTTTC 556
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5772 AGTGGGTTCCAGAGGGAGTCCCAATTAATTCCTTTT ---GATGATGCGATTGACTTGC 5828
FT QY 557 AACAGGGAAGTGTGCGACCAACGCTTTTATATAATCTACCAAGTAGTGTGTTTAGACGTAAA 616
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5829 GACTCATGAAGTTGCGACCATGTGTTTATCAACTTACCAGTAGTGAACATTACGATCTCAA 5888
FT QY 617 GCATTCGTTTCAGATTTTGAGTTGTTAGTATTGATGTAAGCGTTGATATAATTCATT 676
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5889 ACATCTGTTTCCGATTTTGAAGTCATGGGTATTGATGTGATGTAGATATAATTAATTAAT 5948
FT QY 677 CGGTTTACTCGGTTGAAAAACAAAAAATCCAACTGCTAGCTGACCATAGCATTTGTAAC 736
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 5949 TGATTTCCGGCTTTAAAAAATAAAAAATCAACAGTTGAGACCATAGTATCGTAC 6008
FT QY 737 TTTTTCACAAATTTTATTAAGCCCTAGTCATATCATGATGAACGACGTTTTGGATATACT 796
FT DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
FT 6009 TTTTAACTCCAATTACTACAAACATAGCCATATCTTTCTAAAGCGCATGTTGGATATCTT 6068
```


QY 133 TTTTATATCTCGTAGCCAGTTGAAATTTAGTATAGAGGTAATCTGATAGAGTTTGAAA 192
Db 4644 TTTATCTCTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 46383
QY 193 AAAAATTTAACTATAGATATAATTTTCAATTTTCTTACGGCAGTATCATTTTGTGG 252
Db 46382 TGGTATTCGAATACAGCTTTTACTATATTTTCAATATCAAGTTCAATATTTTATTTT 46323
QY 253 AGAATAATTTTCGCACCTTTCAAGACGTTGGTGGCTGATTTTCAATATAATAAATTCGTTT 312
Db 46322 AAAAATCTTTTACACGACACGACTTCTCTTTTACCTTTTATTTTATTTGATGAATTCGATTT 46263
QY 313 TGGTATACCTATTTAAACGTAATTAATAGCAGTTTAAGGATAGCTTCTATTTTTCGACAA 372
Db 46262 TATTATATCTATGAAATTCATTTTAAATATATATCGAAAAATATTTCTTACGCTAAGTTT 46203
QY 373 TCTATCAAAAAAGACGATCTCAATTAACAACGGCTGAACGATGGGAAATATGCAAGTTT 432
Db 46202 CACGAGATACCAAGTTGTTTGTATACGAATAAGGATCTTTATCAAAAATGACCTTTA 46143
QY 433 TATTTGAATCACATAAAACAAATTCAAAAAATCTTGTTCATTTGGTATTTTAGGTACAG 492
Db 46142 GGAATAAT ---ACGACATAATATATCGCTGCTGTATCTTGGATTCCTCGAAAGG 46086
QY 493 AAATAGATAAAATTAATTTATCATTAACGCTCTATATTTCTGTGGAAGAGCTATAGAGT 552
Db 46085 ATTGTTATGATTTGAAACATAACTCGTTAAGGATAATAAAACAAAGATCTCTTACTTCAG 46026
QY 553 TTTCAACAGGAGTGTGCGACACGCTTTTATATAATCTACCAAGTCAGTTTTCAGC 612
Db 46025 AGTTAACTGCTTAACCTGTGATCAAGCTTTTATTAACATACCCCAATGAAATTTTGGTA 45966
QY 613 TAAAGCAATTC-----GTTTCAGATTTTGTAGTTGTTAGGTATTCATGTAAGCGTTG 663
Db 45965 AATACCAATACAGATATTAATGACATTTGAAGCAATGGAGTGTATGTCATGTTA 45906
QY 664 ATATTAATTCATTCGGTTTACTCGTTTGAATAACAAAAATCCAACTGCTAGGTGACC 723
Db 45905 ATGTAGAGGCACTTAGCTTTCATATA TAGGAGAAAGCGAATCCAACTTTTGAAGAT 45846
QY 724 ATAGCATTTGAACCTTTTCCAAATTTTATAGCCCTAGTCATATCATGATGAACGAC 783
Db 45845 ATAGTGTATTAACATATTTCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT 45786
QY 784 TTTTGGATATATCTCGACGCTGAGTCGGTTAAATTTATTTGGTATAGTTCTATTTTGT 843
Db 45785 TTTTGGATATACGGGTGCTATATAGGTTTGTCTCATATGCGCATTTGGCAATTTTTC 45726
QY 844 TAGTTCCAAATTTTCGTAGAGATGGTGACCGGCTATTTTTCGTAGAAACGAGTTGGAC 903
Db 45725 TAGTTCCAAATCAGAAAGATGGTGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC 45666
QY 904 AGAATGGACGATATTTATCATTTCAAGTTTCGATTCGATGTATGTTGATGCTGAGGAC 963
Db 45665 GTAATGGTAGGATTTTATAGTTCTATAAAATTCAGATCAATGCGAGTAGATGCAACAAA 45606
QY 964 GCAAAAAGACTGTCTAGCCAAACACAGATGCAAGGTTGGTATGTTTAAATGGGAA 1023
Db 45605 TTAAGAAAGATTTATAGTTCAATCAATTAAGACAGGG-----CTAATGTTTATAGTTAGA 45551
QY 1024 AAACCATCTAGAAATTTACTCAATTTGGACATTTTCATAGCCAAAAACAAGTTTTCAGCAGT 1083
Db 45550 CGATGATCTAGAAATTAATAATAGGAAATTTATTCG-AAAAACAGCATAGATAGT 45492
QY 1084 TACCAAGTTTATTAATGTTTAAATTTGGCGATATGAGTCTAGTTGGTACAGTCCACCTA 1143
Db 45491 TGCCTCAATTTCTATAATGTTTAAAGGGTGTATGAGTTTAGTAGGAACAACGCCCTCCCA 45432
QY 1144 CAGTTGATGAATTTGAAAAATATATCTCTGCTGCTCAAGAGAGAGGATTTAGTTTAAACCG 1203
Db 45431 CAGTTGATGAATGAAAAAGTATATTTCAACGACAGAGCGAGGCTTAGTTTAAAGCCAG 45372
QY 1204 GGAATFACAGGTCCTCGGCAGGTTAGTGGTCGTAGTAAATATCACAGACTTCGACGAGTAG 1263

Db 45371 GAAATCACTGGTTTGTGGCAAAATATCTGTAGAAAATAATATTTACTGATTTTGTGAAATCG 45312
QY 1264 TTCGGTTTGGACTTAGCATACATTCGATTAATTCGACTATCTGTGTAGATATAATAAATTTTAT 1323
Db 45311 TAAAGTTAGATGTTCAATATATCAATGATGCTATTTTGGTCAGATATTAAGATTATTC 45252
QY 1324 TAAAGACAGTGAAGTTGTTATTTGTGTAGAGAGGGAAGTAAGTAAAGTAT----- 1373
Db 45251 TCCTAAACACTAAAGGTAGTTTCTACGCGGACAGGAGCTAAGTAAAGGTAAAGGTTTGAAG 45192
QY 1374 -----ATGAAAGCTTTTGGTGGTCTTTCAGGGGACATTTGACTCACTCTGTATTT 1426
Db 45191 GAATATATTAAGAAATTTGTCTGGTTGGTTCAAGTGGTGGTCACTAGCACACTTTGAACCT 45132
QY 1427 GTTAAAAACCGTTTGGAGGAAGAAACGTTTGGGTAAACATTTGATAAAGAGAGATGC 1486
Db 45131 TTTGAAACCCATTTGGGAAAGAAAGATAGGTTTGGGTAACTTTGATTAAGAGATGC 45072
QY 1487 AAGAGCTTTTGAAGAAATGAATAAATGATTCATGTTTTCCTTCAACAAATTCGCAATCT 1546
Db 45071 TAGGATATTTCTAAGAGAGAGATTTGTATATCAATGCTTCTTTCACAAACCCGTAATGT 45012
QY 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAACCCAGA 1606
Db 45011 CAAAACCTTGGTAAAAAATACTATTTCTAGCTTTTAAAGTCTTGAAGAAAGAACCCAGA 44952
QY 1607 TGTATATTTATCTCTGGTGGCGGCTTGTGCTCCCTTCTTTTACATCGGAAACATATT 1666
Db 44951 TGTATCATATCATCTGTCGCTGTAGCAGTACCATCTCTTTATATTTGTTAGTATTT 44892
QY 1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATTCGAGTTAATAAATCTACATTAACCTG 1726
Db 44891 TGGTTGTAAGACCGTTTACATAGAGGTTTTCGACAGGATGATAAACCACTTTCACAGG 44832
QY 1727 AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAGTGGGAGAAATGAAGAGGT 1786
Db 44831 AAAATTAGTGTATCTCTGTAAACAGATAATTTATTTGTTTCAGTGGGAGAAATGAAGAGGT 44772
QY 1787 ATATCCTAAATCTATTAATCTTGGGAGTATTTTATGATTTTGTAAACAGTACGAACT 1846
Db 44771 TTATCCTTAAGCAATTAATTTAGGAGGATTTTATGATTTTGTACAGTGGGACA 44712
QY 1847 CATGAACAAACAGTTTAAATTCGATTCGATAAAGAGATTCATTTTATGAAAAAATGGAAGT 1906
Db 44711 CATGAACAGCAGTTTCAACCGCTTTATTAAGAGATTTGATAGATTAAGAGGACAGGTCT 44652
QY 1907 ATAAACGACGAAATATTTATTTCAACAGGATTTCTGACTATATTTCCAGAAATTTGCAAG 1966
Db 44651 ATTGATCAAGAAGTGTTCATTTCAACCGGTTTACTCAGACTTTGAAACCTCAGAAATTTGTCAG 44592
QY 1967 TATAAAAAATTTCTCAGTTACAAAGAAATGCAACAAATATATTAACAAATCAGAACTAGTT 2026
Db 44591 TGGTCAAAATTTCTCATATGATGATGAATGAACTTTTATGATGATGATGATGATGATGAT 44532
QY 2027 ATTTGCCACGAGGCGCCGCTACTTTTATGAAATTCATTTCCAAAGGAAAAAACAATTA 2086
Db 44531 ATCACAATGCGCGCCAGCGCGTTTATGATGCGAGTTTCTAAAGGAAAAAACAATTT 44472
QY 2087 TTGTTTCTAGACAAAAAAGTATGGTGAAACATGTAAATGATTCATCAAGTATAGATTGTA 2146
Db 44471 GTGGTCTCAGACAAAGACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGACTTCTGT 44412
QY 2147 AGAAGATTTTACAAGATAATAATATTTATTTATAGAAATATAGATGATTTGTTTGA 2206
Db 44411 AATAAGGTAAAAACAATGTATATTTTATGATTCGTTGTAGATATTTGAAGGTTTACAAAT 44352
QY 2207 AAAATTAATGAGTTTCTTAAGCAAACTAACTTTTACATCAAAATAATAATTTT---TTTTGT 2263
Db 44351 GTAGTCTATAGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44292
QY 2264 GAAAGATTTAAACAAATAGTTTGAAGAAATTTAATGAGGATCAAGAAATTAATAATAATAA 2323

Db 44291 GAAGAAATTAAAGTAAATATTAAAGGAGTTGCGATGAAATCAATAAAAACTCTTTATT 44232
Qy 2324 AGATGCATATTTGATAATGGCTTATCATATTTTCTCAGATTTTACTGGAGAGGATAC 2383
Db 44231 TTATATTGCAATATTTTATAGTAAATTTTAAATCACTGGGTTTAGCGAGGAAACTC 44172
Qy 2384 AGATATTATCATCTTCTCTCAGGAGATGACACCATTTAGTTCCTTCAGAAATCCTGTAT 2443
Db 44171 AGCTTACAAATAGTG-----ATGTTAGTTGCAA 44143
Qy 2444 AATTATTAAATATCTCAGAAATTATATGTTGAAATTTTACAAAGATGACAAAAATAT 2503
Db 44142 TTTTACTGTGGGAATAAAATTTTTTATAGATGCTTTTATTTTGAAGAAGAAAACTCG 44083
Qy 2504 AAAGAAATAGGATATATGAACGAGTTAAATGTTTACAGATTTATTTCTTAATATACGAA 2563
Db 44082 TGATCATCTTTTATTTATTTATTCGCGACCAATTTTGAATTTATTTCT----- 44038
Qy 2564 AAAACTATTGATAATGTAATGTTTGTAGAAATTTTATTAAGAATGTATCGAGCTTTTGAATAC 2623
Db 44037 ----TTGTTCTAAGGTTACTTTTATATTAATTAATTTTCTTAACTTTTATTTTCTAGCAATTAAGGAC 43982
Qy 2624 TATTTACAAAGATTTGTTTATTTGATAGAAATAAAAAACATGGTCTAAGAAATAAGATTTG 2683
Db 43981 ATCTCTCAAABAAAGCTTTCTCTATAATATAGATCGCGTATTTTGGGAGTTCTATTA 43922
Qy 2684 GTTCTAATGTTGGTTTCGCTCCACAGATTTTGGGCAATTCCTTTATCAAAATGAAACG 2743
Db 43921 AATCAAAATTTTGTGAATTAAGATTTAATAGAAATTAAGTATGTCAATTTTATAGGGAT 43862
Qy 2744 AAACAGCTTATTTATTAAGTAATCTAAATCTCCAGATGAATTTATACAGACAAATTA 2803
Db 43861 GGACAAATTTATCGAGAAGTGA-CCTTAGGTTTGGTCATCTCACTTTATTCATAATTT 43803
Qy 2804 TAGAAAAATATGAATTTTCAATAGATTAATCTAAATATGAA-----ATTAAAGATATATA 2860
Db 43802 TTTTGCCTCACTATTTCTGTATATTTGTACTCAATATTAACGACTAAAGCCTGTTGT 43743
Qy 2861 AGTGAAAAAATCAACATCTTCTCCTATGTCCTTTTACAGATGATCTTATGATGAATGTC 2920
Db 43742 GATGGTTTATTTTAAATTAATTTATTTGTTACCAATATACTTTTCAAGGACAGG 43683
Qy 2921 TAAATGCAAGAAATTTAGGTTTTTATTTGCTAGAAAGTTAAAAATGAAATTAATCTA 2980
Db 43682 GTATTATATCGTAATTTTATTTATTTGTACTCAATTTATGTGCAAGAAATAGCTTAATAA 43623
Qy 2981 AA----TTTAAAGAAATTAATTAATAAAAAATAAAATAGTTGATTTGTGAGAGTAATGTA 3036
Db 43622 AAGAGTATTTATGAAATTTAGCACCTATGTACAAATTTTTTATATAGTATTTACCTTTTT 43563
Qy 3037 TGTTTAAATATTTAAATATGACCCGGGAATATTTTATTTTATTTTATGATCTTCTGGTTGATTA 3096
Db 43562 GAGTCTCAAAATTTTTTAAATCAAAATTTTGTTCMAAAATTTAGATGTTCTTTTAAACAGG 43503
Qy 3097 TTTTATTTCCAGACCAAGATGATGATTTTATTTTAAATTTTATGAAATTAATTTATTTTC 3156
Db 43502 TAGATTACATATGCTCATTTTCAACCTTTGATAGTGGTTTAACTCCTTTTGGAAATAGTTT 43443
Qy 3157 ATATAAAATTTTGAATACTAAGCTAAATATTTTAAAAAAATGAAATTTTATTTTATTTAT 3216
Db 43442 TAAGGAAACAAGTGTCTTATTTGATATATAGCTACTCTATGTTATTTGATATGATGTTGT 43383
Qy 3217 GGTCTATATTATGTTTGTTCAGTGTGACCAAGTATGTTTGTGAAATAAAATTTTGA 3276
Db 43382 AGTACTTACCATGTTTGTATGATAATCTATATATCTATAGTAAAGA-----TAA 43330
Qy 3277 GATTATTTGAGATTTTACTGCTCCCAATAATTTGGATTTATTTGCAATATGATTTATTAAT 3336
Db 43329 TCATAATTTGAACTTCAACTACTCTCTATTTATTAATGTTCTATATATATTTTACTGAAAGTT 43270
Qy 3337 TGTATTTCATTTATAAATATGATTTATAAAAAATTTAAAAATAGTATCTTTTATGTTTTT 3396
Db 43269 TTTATCCAGTGTGTAATGAATATTTAGTTGGCTAGTTTTTGGTAAAAATATTTTGTGATG 43210

Qy 3397 TAGTTTTTATAGGTATATCTGCATCTGTATATTATTCAAAATGGAAAGATATTGTATTTT 3456
Db 43209 GTATCGAACCTATAAAAAAGGAATTTACTATTGTGAATAAT--ATATGACATATTTGCTC 43152
Qy 3457 TAGACAGACACCTTTATAGGACTAGACTATCTTATTAACAGCGCTCAAAACAAGGTTGGTTG 3516
Db 43151 TGATATGGCAGGAGGTAAAGGAAGAAATGATACCTTAAAGTTATACATATTATTGTTGTTT 43092
Qy 3517 GCTTTATGAACACTATCCCTPACGTTAAATACCACTACAAATATATAGTTTCAATTCOGTTAATCT 3576
Db 43091 GGAGGAAATCCCTTACCAGATAATTTAAAGAAATATATAAAAACTTGGAGAGAAACAATGT 43032
Qy 3577 TTGCACCTATAAAAATAAATGCAACAATTTTTTTCTTGTGCTCTGCTTTTATACCGA 3636
Db 43031 CCGGATTTATGAAATTTTGAATGGAATGAGCATAAATTTATGATGT----- 42988
Qy 3637 TCTATTTTAAGTGGATCGAGAAATTTGGTAGTTTATCGCTPAGCAATATTAATTTATATGCTTGT 3696
Db 42987 --TAGTAAAAATGTTTTTATGAGAGAGCAATATACHTAAGAAATTTTGGCTTATGTTTCT 42930
Qy 3697 TATGGAGATATATAGG-TGGAAAAATTTTCTTGGATAAAAAGCTAAATPAGTAATATTTGTA 3755
Db 42929 GACTATGCAAGATTTGGATATTTATATACTTTATGGGGGTTCTATCTAGATACCTGATGTG 42870
Qy 3756 ATACTACTTATTTTAAATACTGAATTCGCTTTACCATGAAATTTTGGCTGTTTATAAT 3815
Db 42869 GAGCTTTTAAAAAGTTTAGATCCTTTTGAGGATTCATGAGTGTCTTCTAGCAAGGAGATT 42810
Qy 3816 TCTAGAGAAATCAAGTACGAACTAGATTTTATTTATCAAGGAAGTATTGTATAAAGTA 3875
Db 42809 AGTTGATGTGAATACAGGATTAATTAATTTGGCGCTGTTAAAGGACATCATCTTTTAAAA 42750
Qy 3876 TTAGAAAAAATATTTTATTTTGGATATGGAATATCCGAATATTTCAGTTACGGGAACCTGG 3935
Db 42749 TCAATATATGCTATATATACAAAAAGTGATTAACCTCTCTTAATAAGACATGTTAGAG 42690
Qy 3936 CTGGAAAGTCAATTCAGGCTATATATCAATTTTTTATTAATCAGGAATPAGTTGGGTTGAT 3995
Db 42689 GTTACAACTAATTTTATGATAAAACAGAGGCTTTAAGAAAT--AAGAAATATTATTCAAAAGA 42632
Qy 3996 TTACTGATGTTTCTTTTTTTTTTATGTTATATAAAAAAGTTATGAGTAAATGGGAAACA 4055
Db 42631 TTGATGATATAACAATAATATCCGAGAAATTTTATTAATCCAAAGAAATTTATTAACAGTA 42572
Qy 4056 GCATATTTTATTTTATCATCATTTAGCCATATTTTTCATATATGAAACAATAGATCCGAT 4115
Db 42571 AGGTTGATTTCTGACTAGTGTACCTATTTCTATACATCATTTACGAAGGAGTTGGAAAA 42512
Qy 4116 ATTATTATATTAGTACTATTTCTTTTCTCAATAGGATTTTGGAAATATATAAATTTTAAA 4175
Db 42511 GTTCTTCAATTTATTTTCAAGATTTCTTAAAGATTTAGAGTAAGGCTCATTAATTTGATTTTAT 42452
Qy 4176 AAGGATATGAGACAAAAAAATGAATTTAAATTTTCAAGTTATTGTACCAATTTTATAATGT 4235
Db 42451 TTGATATAGTACTTTATAGAAATGCTTCTAAGGTTCTTAAAGTTTAAAGAAATAGTTATAT 42392
Qy 4236 CCAAGATTTATCTGATAAATGTTATTAACAGTATTTTAAACCAACATATATACTAATTTAGA 4295
Db 42391 TGTAAATAGAAGCGAATGTTATATATTTTACTCTACATTTTAAACAGGCTACCGA 42332
Qy 4296 GGTATTCTCGTAAATGATGGAAGTACTGATGATTTCTGAGAAAAATTTGCTTTAAACTATAT 4355
Db 42331 TTATCCTATC-TGTATGATTTCTCTATGTAATCAGACAAATAAAAAATTTTATATGTTGAT 42273
Qy 4356 GAAGACGATGGAAGAATTAATAATATATCAAGAAAAATTAATGGCGGTCTAGCAGATGCTCG 4415
Db 42272 TGTGATGATGGTCTGAGGACAGTACTAAGGAAATTTGTAAGTAATTTATATAAAGAGAA 42213
Qy 4416 AAATTTCCGACTAGAACATGCAACAGGTAATAATATTTGCTTTTGTCCGATTTCTGATGACTA 4475
Db 42212 TAAAGTTAGTATTTGTATCTGTA---TAAACGTAATGGCGGTAAAGCATTCAGCCTATAA 42156

QY 4476 TATGAAGTTGCAATGTTTCGAGAGAAATGCATGATATATACTAGTATATATCCGATAT 4535
DB 42155 TCTGCAATAGATATATGCAACCAAGTATATCATGTATCTGTAGATAGTATG 42096
QY 4536 AGCAGAGATAGATTTTTTGTAGTAGACGAAACGGGTATACAAAGAAAAAGAAATAG 4595
DB 42095 GTTATTAGAGATGATGAGTTGAAATTTATTTAAAGATTTAGAGACCTTAAACGTTGCTTAA 42036
QY 4596 TAAATTTTCAATGCTTAAACAGAGAGAGACGCTGAAAGAAATTTTTTGTGAGGATCTTAATAT 4655
DB 42035 TAGATATGTAGGCTT-----TGTTTATCCTCGATATTCATTTAAACCAAGGTAAATATG 41982
QY 4656 AGAAATATATGTTTGGTGCAGCTTTATTCACGAGATATATAAAGATATAAAATTCCTCA 4715
DB 41981 GTTGAATCCCAAAATATTAGAGTTTAAATTCCTGATTTAAATATAAATATCATCTTAA 41922
QY 4716 AATTAATATAGAAATGTTGGTGAGATTTGCTTTTAAATTTGGAGGTCTTGAACAAATGT 4775
DB 41921 AATTCGAATGTTATGTTTAAATTAATAA--TGCTTATTTAGTAGATTTGCAATTCCTTGT 41864
QY 4776 AACACGTGTAGTGTGATCTAGAGAAATATTATTAATATGATTCATTCGTAAACAGTTTC 4835
DB 41863 TTCGAAGGGAGAAATTCCTTATCAGAGAAATAATGATATATATTTTATCAAAAAGGGA 41804
QY 4836 GCTTATTAATCGAAATTTCTCTATAAATAATATGATTTAGTTCACAAAGTTGAGAAATTA 4895
DB 41803 TACTTTTGTCCCAAAATAGGAAATCTATTG-----TTTTCGACTACTTAGAAGATGGTTT 41748
QY 4896 CCCCTTTAAGTTAAAGAGAGTTTGTAGTCATTTTGTGATGCAAAAGTTATTAAAGAA 4955
DB 41747 GACTTCTAATTTTTTAAATTTATGGAGAAATAATTTTCAAGGGACTATTTTTTCATTAGA 41688
QY 4956 GGTTAATGTTTAAACAAATGTTTCAACAGATTTGTTGGATAATGAGTTCTTCCCAAT 5015
DB 41687 AAATTCATATATGATGTAATGCTTTTCCCTAAATATTTTGTATCCTGCTGCTGCTCAGC--T 41630
QY 5016 ATTGAGTCTTATCGAAAGAAATACGTAGATATCCATTTATTAAGCGAAAGATATTT 5075
DB 41629 ATAAAGATAAAATGAACATACAGCTTTGAAATGACTTTTAGGGGTAAATACCTACA 41570
QY 5076 ATCAAGAAAGCATTTAGTTAGTGTGTTATTTGATGAAATTTTCGCTCAACTATATGTAAT 5135
DB 41569 CTAAG-----AGTGAAGAGCTGGGTGGAAATTTTTTGTGGGTTTTAGCTATCTT 41519
QY 5136 GTTATATAAGAAATTTCAAAAGCAGTAGAGGTAAAAATGGATATAATTTAGTTTATGTT 5195
DB 41518 TGGAAGGTAGCAAGTTTAAAGAGCGAGTAAAGAAATGGAAATTAATTTCAATATAGTG 41459
QY 5196 CCAGTTTATATGTAGATAAATATTTAAGTAGTGTGTATAGAAAGCATTTAATCAAAAT 5255
DB 41458 CCTGTATATATGGAGAAATATATATGGAGATGCTAGATAGTATTTCTTGAGCAAACT 41399
QY 5256 TATAAAATATAGAAATATTATGATAGATGATGCTCTGTAGATGATTTCTGTAAATA 5315
DB 41398 TATCAGAAATTTAGAGATTTATCATATTTGATGATGCTCTAGTGTATGCGACAGGGATAT 41339
QY 5316 TGCAAGGAATATGCAGAAAGATATAAGAGTAAAAATTTTTTTCACATAATCATAGTGA 5375
DB 41338 TCGGAAAGTATTTTTTGGAGAGCAGGCGAATAAATAATTTCTATCAAGAAACAGAGGC 41279
QY 5376 GTATCAAAATGCTAGAAATCATGGAAATAAGCGGAGTACAGCTGAAATATATTATGTTTGT 5435
DB 41278 CAATCAGTTGCAAGTAAACAGGAGTACTTAGGTGTACAGGAGATTTGATTTCTTTTAA 41219
QY 5436 GACTCTGATGATGTTGTTGATAGTATAGATTAGTAAAAATTTATATTTTAAATATATAAA 5495
DB 41218 GACAGTGTATGATGTTTATCTTCTTATTTCAATTTGAGGTTTATGTAACAATACAGAAAGCT 41159
QY 5496 AGTGAAGTATTTATCTGTTGTTGTTGACGCTACTTTTTCAGAAAAATATAAATTTT 5555
DB 41158 ACCAATGCGGATATTTGTTTAACTAGTATGTTTAACTTCAACAATATCTACTAGT 41099
QY 5556 GAAGTGAATAATCCAAATATTGATTTTGAAGCAATTAATACCGTGCAGGACATGGAGAA 5615

DB 41098 ATTAAATTCACAGTATTTTAAAGAGATTAAACCTTTTACTTTTAGAGGTGCGTGTGAAGAA 41039
QY 5616 AAAAATTTTATGAATTTGTATATAATAATATTTTTTCTACTCTCTGTTTGTAAACTATAT 5675
DB 41038 ATGTATTTATGGGAAACCTTTATGGAGTTTCCCGTTAGCAA-----AATTTATCCGAGAA 40984
QY 5676 AAGAAAGATACATAACAGATCTTTTTCAAGAGAAATCAATGGTTAGGAGAAAGATTTACTTT 5735
DB 40983 GTAAATTTATGAGTAATCCATATCCAGAAAGGAAAAATTCATGAAGATATGGAATAACAAT 40924
QY 5736 TTTAATCTGCATTTTAAAGAAATATAGATAGAGTTAGTTATTTGACTGCAACATCTTTAT 5795
DB 40923 TTAAGCTTATATCTTTCGCCCTTCAAAAATAGCAGTTTGTGATATTTGCTACTGCAAGTGTGT 40864
QY 5796 TTTTATAGCAGAGGTATATAAGTACAGTAATTTCTTTTAAAGAGAGGTGTGTTTTTGCAG 5855
DB 40863 ATTTTAGCGATATAGTACTACTAGGACAAATTCATCAGNGAATGT-TGTATTTTTTTT 40805
QY 5856 TTGGAATAATTTGCAAAAACAAAGTATAGTATTTGTTTAAAGCAAAATATATGGTGAGGATTTT 5915
DB 40804 GAGGCAATTCAAAATAATATTGTTTATAAATCTAAATTTTCCACATAATACATCTTTTA 40745
QY 5916 GACGTATCAATTTTAAAGATCTATACGTTGGCGAAGTATTTTATATAGCTTACTAATG 5975
DB 40744 ATCTCTGCTGTAATATATAATGAAGTATTTGGTGGAAATAGATATATGTGGGAAGATGAT 40685
QY 5976 TTTAAATACGCAAAACAGCTCTATTTTTCACAAAATTTTTTAAATTTTATAGAAATCTTTTATA 6035
DB 40684 GATTTTAAATTTATGATGATACAGTTGATTTATACAGAAAAAATAATATAGAAATATTTTAA 40625
QY 6036 AAATATTTTAACTTGTTTAAAGTATCTAACAAAAATCTTTGCTCTAAAAATTTTTTGT 6095
DB 40624 ACTATCTCTTTTAAATATPAGGATTAGTGTGAAGGAGAGGTTTAAATAATATATTTATTT 40565
QY 6096 ATAAGAAATGTTTCGAACAAAAGTTTTTAAAAAAATTTATGTTTATATAGGAAGATATC 6155
DB 40564 TCCTCTATAAGATATTTTCACTATTTGTAAGAAAAATA-----TATAAATTTGAGGTTATC 40512
QY 6156 ATGATACTATTTAGTAAAAATTTCTATATTTGTCACCTATATATAATGTAGAAAAATTTTATA 6215
DB 40511 AAGGA---ATGGATTAAGTTAGTATCATTTATCCAGTTTATTAACGTTCAATCTTTCTTA 40456
QY 6216 TCTAAATGTATAGATAGCATTTGTAATCAGAACCTTACAAACATATATAGAGATTTCTCTGGTG 6275
DB 40455 AACGAGTGTATAGAAAGTTTGTAGCTCAAACTTATTCAAAATTTAGAAATTTATCTAGTT 40396
QY 6276 AATGACGTAGTACGATTAATTCGGAAGAAATTTGTTTATAGCATATGCGAAGAAAGATAGT 6335
DB 40395 AACGATGCTCTACAGATAAATTCGGGGATATCTGTGATTTATTCAGAAATAGATGG 40336
QY 6336 CGCAATTCGTTATTTTAAAGAGAACGCGGGCTATCAGATGCCCGTAAATTTATGCGATA 6395
DB 40335 CGTATATTCGTTTTCAAAAAATAAGTGGTGTATCAGATGCTCGTAATTTATGGAATTT 40276
QY 6396 AGTCGCGCAAGGTGACTACTTACTTTTATAGACTCAGATGATTTTATTTATTCATTCGAG 6455
DB 40275 AGTAGAGTACTGGGATTTATTTATTTACTTGTTCGGATGATTTCTATATTAAGAA 40216
QY 6456 TT---CATCCAACGTTTACACGAACAAATTCAGAGAGAAATGCCCCTTTGTGGCAGTTGCT 6512
DB 40215 GATGCAATTCGAGAGAAATGTTTGAATTTAGCGAGAAATATAATTCAGAAATTTGTTTTAGGG 40156
QY 6513 GGTATG-----ATAGGAGTATGCTTCGGGCAATTTCTTAACAGCAGAGCCGCTTCT 6566
DB 40155 TGTATTGTGAAAAAAGAGAAACAATATTAATAATCTCGTTTTTATAGAAATGAAATGATA 40096
QY 6567 ACAAATCAGGCTGTTCTGAGCGCAGGAATGTTTGTAAAAAGCTGCTAGAGCGCGATGCT 6626
DB 40095 GAAACTATAGTCAGTACAGCAATACAAATATTTTATAA---TTATGATGCCCTATAGA 40039
QY 6627 CATCGCTTTGTGGCCTGTAAATACTCTATAAAAAAGAACTATTTTGAAGATTTTTCGA 6686

Db 40038 GCTATTTTCTAGTACACATATAAATTTATATAAGAGAGAAATTTATTTCTACGCTGTC 39979
Qy 6687 TTTCGAAAGGTAGATTCATGAAGATGATCTTCACTTATCCCTTGCTCTATGAGTTA 6746
Db 39978 TATCCAGTAGGGAAGTTTACACGAAGATGAATTTTAAACCTATAAATTTATTTTGAAGCC 39919
Qy 6747 GAAAAAGTTGCAATAGTTTAAAGGAGTGTCTGTACTATTATTGTGACCGAGAAATAGTATC 6806
Db 39918 AAAATATAATTTCTTTAGATATAATACCTATATCGTATTAGAGAAAATAGCATA 39859
Qy 6807 ACAACTTCTAGCATGACTGACCATCGCTTCCTCACTGCTTCTGGAATTTCA 6856
Db 39858 ATGACTGTTCTTACATATTAAAGGCTACATCTGTGAGAAGCACTTAA 39809

RESULT 6

ID ABK90550
AC ABK90550 standard; DNA; 6865 BP.
AC ABK90550;
XX
XX 15-NOV-2002 (first entry)
DE Betal,3-galactose transferase DNA #2.
XX
KW Betal,3-galactose transferase; gene; ds; galactose; N-acetylglucosamine;
KW betal,3-bond; galactose-containing glucide; uridine-5'-diphosphate;
KW infection.
XX

Streptococcus agalactiae.

Key Location/Qualifiers
CDS 617..1792
FT /*tag= a
FT /product= "Betal,3-galactose transferase #2"
FT 1816..2262
FT /*tag= b
FT /product= "Betal,3-galactose transferase #3"
FT /note= "No stop codon given"
FT 2265..2747
FT /*tag= c
FT /product= "Betal,3-galactose transferase #4"
FT 2843..3979
FT /*tag= d
FT /product= "Betal,3-galactose transferase #5"
FT /partial
FT /note= "No stop codon given"
FT 3982..4956
FT /*tag= e
FT /product= "Betal,3-galactose transferase #6"
FT 5009..5950
FT /*tag= f
FT /product= "Betal,3-galactose transferase #7"

JP2002199885-A.

16-JUL-2002.

05-JAN-2001; 2001JP-00000392.

05-JAN-2001; 2001JP-00000392.

(KYOW) KYOWA HAKKO KOGYO KK.

WPI; 2002-612563/66.

P-PSDB; ABG31703, ABG31704, ABG31705, ABG31706, ABG31707, ABG31708.

Betal,3-galactose transferase and a DNA encoding the enzyme, a protein having betal,3-galactose transferase activity, a transformant, preparation of a protein, and preparation of a galactose-containing glucide.

Example 1; Page 15-22; 24pp; Japanese.

The invention relates to a protein having betal,3-galactose transferase activity, derived from a microbe having an activity of transferring galactose to N-acetylglucosamine by betal,3-bond. The protein can be used in a method for the preparation of a protein having betal,3-galactose transferase activity in which the above transformant is cultured in a medium to form and accumulate a protein having betal,3-galactose transferase activity, and in a method for the preparation of a galactose-containing glucide in which a culture liquid of the above transformant or a treated product of the culture liquid is used as an enzyme source, the enzyme source uridine-5'-diphosphate galactose and a receptor glucide are made to be present in an aqueous medium, and a galactose-containing glucide is formed and accumulated in the aqueous medium. The galactose-containing glucide is used as a candidate for infection preventors. This sequence represents DNA encoding betal,3-galactose transferase proteins of the invention

SQ Sequence 6865 BP; 2453 A; 847 C; 1185 G; 2380 T; 0 U; 0 Other;

Query Match 10.4%; Score 727.8; DB 6; Length 6865;
Best Local Similarity 59.2%; Pred. No. 6.9e-89;
Matches 1389; Conservative 0; Mismatches 917; Indels 41; Gaps 7;

Qy 16 TGGCATTATTGTATGATAGACAGTTGCCAATTTCTGCAATCTTAAACAAGTCATATACCAA 75
Db 438 TAGCGATGATTCAACACCGTTGTGGTTATTTTCTGCAAGTTGACATTAACATTAATTA 497
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATATGATGATGTTTCATTTTTCGAT 132
Db 498 CTCCTCACTTTAAAGCAATAAGATTATTGTTGTTCTTATGATACATATTATGTTCT 557
Qy 133 TTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGTAATCTGATAGAGTTGAAA 192
Db 558 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 617
Qy 193 AAACATTTAATCTAGTATAATATTGCAATTTTCTTACGGCAGTATCATTTTGTGG 252
Db 618 TGGTATTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTCATTTTATTTTCTA 677
Qy 253 AGAATAATTTGCGACCTTCAAGACGCTGGTCCGCTGATTTTACATTAATAAATCTGTT 312
Db 678 AAACTCTTTTACACGACGACTTCCCTTTTACTTTTATGCTATGATTCGATTT 737
Qy 313 TGGTATACCTATTAAAGTAATTTAAAGCAGTTTAAAGTAGCTTTCTATTTTCGACAA 372
Db 738 TATTATATCTATTGAAATTCATTTTAAATAATTTATCGAAATATTTCTTACGCTAAGTTT 797
Qy 373 TCTATCAAAAAGACGATTTCTAATTACACGCTGAACGATGGGAAATATGCAAGTTT 432
Db 798 CACGAGATACCAAGTTGTTTGTATACGAATTAAGGATTTCTTATCAAAAATGACCTTTA 857
Qy 433 TATTTGAATCACATAAAACAAATTTCAAAAAATCTTGTTCATTTGGTAGTTTATAGTACAG 492
Db 858 GGAATAAAT---ACGACCAATAATTATATCGCTGCTGTAATCTTGGATCTCTCGAAAGG 914
Qy 493 AAATAGATAAAATTAATTTATCAATTTACCGCTCTATTATTCTGTGGGAAGAGCTATAGAGT 552
Db 915 ATTGTTATGATTTGAAAACATAAATCGTTAAGGATAATAAAACAAAGATGCTTCTTACTT 974
Qy 553 TTTTCAACAGGGAAGTGTGACACGCTCTTTTAAATCTTACCAAGTGTCTTTTAGACG 612
Db 975 AGTTAACTGCTTAACTGTTGATCAAGCTTTTATTAACATACCATTTGAATTTTTCGTA 1034
Qy 613 TAAAGCAATTC-----GTTTTCAGATTTTGTAGTTTGTAGTTTGTAGTGAAGCGTTG 663
Db 1035 AATACCAATAACAAGATAATTATTAAATGACATCGAACCAATGGAGTGATTTGTCAATCTTA 1094
Qy 664 ATATTAAATCAATTCGGTTTATCTGCGTTGAAAACAAAAAAATCCAACTGCTAGGTGACC 723
Db 1095 ATGTAGAGGCACCTAGCTTTTGAATAATATAGGAGAAAAGCGAACTTCTTGAAGGAT 1154
Qy 724 ATAGCATTTGTAACCTTTTTCACAAATTTTATTAAGCCTAGTCATATCATGATGAACGAC 783


```
Db 6016 TAAAGTTAGATGTTCAATATATATCAATGAATGCTCTATTTTGGTCAGATATTAAAGATTATTC 6075
Qy 1324 TAAAGACAGTGAAGATCTGATTTGTTGAGAGGGAAGTAAGTAAAGTA----- 1372
Db 6076 TCCTAACACATAAAGGTAGTTTACTCGGACAGAGGACTAAGTAAAGGTAAAGGTTTGAAG 6135
Qy 1373 -----TATGAAGTTTGGTGGTCTTCAGGGGACATTTGACTCATTGTATTT 1426
Db 6136 GAATATAATGAATAATTTCTGCTGGTTGAGTCAAGTGGTGCATCTAGCACACTTGAACCT 6195
Qy 1427 GTTAAACCGTTTTCGAAGGAAGAAGACGTTTGGGTAAACATTTGATAAAGAGAGATGC 1486
Db 6196 TTTGAACCCATTTGGGAAGAAGAGATAGGTTTGGGTAAACCTTTGATAAAGAGATGC 6255
Qy 1487 AAGAAGTCTTTTGAAGAATGAATAATCTATCCATGTTTACTTTCACAAACAAATCGCAATCT 1546
Db 6256 TAGGAGTATCTAAGAGAAGAGATGTTATCATTTGCTTCTTCCAAACAAACCGTAATGT 6315
Qy 1547 CATTAATTTAGTGAATAATCTTCTTAGCTTTCAAAATTTTACGTGATGAGAACCAGA 1606
Db 6316 CAAAACCTTGGTAAATAATCTATTTCTAGCTTTTAAAGGTCCTTGAAGAAAGAACACAGA 6375
Qy 1607 TGTATTATTTATCTGCTGGCGGTGCTGCTCCCTTCTTTTACATCGGAACACTATT 1666
Db 6376 TGTATCATATCATCTGCTGCGCTGTAGCAGTACCATTTCTTTATATTGTAAGTTATT 6435
Qy 1667 TGGAGCAAGACGATTTATTTAAGTATTTGATCGAGTTAATAATCTATCACTAACTGG 1726
Db 6436 TGGTTGTAAGACCGTTTATATAGAGTTTTCGACAGGATAGATAAACCACTTTCACAGG 6495
Qy 1727 AAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAAGTGGGAAGAAATGAAGAAGT 1786
Db 6496 AAAATTAGTGTATCTGTAACAGATAAATTTATTTGTTTCAAGTGGGAAGAAATGAAGAAAGT 6555
Qy 1787 ATATCTTAAATCTATTTAATCTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGAAT 1846
Db 6556 TTATCTTAAAGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTACAGTGGGACA 6615
Qy 1847 CATGAACACAGTTTAAATCGATTGATAAAGAGATTTGATTTTATGAAAAAATGGAAGT 1906
Db 6616 CATGACAGCAGTTCAACCGTCTTATTAAGAGTTGATAGATTAAAGGACAGGTCT 6675
Qy 1907 ATAACCGACGAATAATTTATTTCAACAGATATTTCTGACTATTTTCCAGATATTTGCAAG 1966
Db 6676 ATTGATCAAGAAGTGTTCATTTCAACAGGTTTACTCAGACTTCGAACCTCAGAAATTTGCAG 6735
Qy 1967 TATAAAAAATTTCTCAGTTACAAAGAAATGGAACAATATATTAACAATCAGAAGTAGTT 2026
Db 6736 TGGTCAAAAATTTCTCTCATATGATGATGAATGAATCTTTACATGAAGAAGAGCTGAGATTGTT 6795
Qy 2027 ATTTGCCAGGAGCCCGCTACTTTTATGAATTCATTATCCAAAGGAAAAAACAATTA 2086
Db 6796 ATCACAATGCGGCCGACGAGGTTTATGTCAGTTATTTCTTAGGGAATATACAGTT 6855
Qy 2087 TTGTTTCTAGACAAAAAAGATGATGGTGAACATGTAATGATCATCAATGATAGAGTTTGTGA 2146
Db 6856 GTTGTCTCTAGAGAAAGACAGTTTGTGTAACATATCAATGATCATCAATATCAATTTTA 6915
Qy 2147 AGAAGAAATTTACAAGATAATAATATTTATTTATAGAAAATATAGATGATTTGTTGAA 2206
Db 6916 AAAAAATTTGCCACCTGTATCCCTTGGCTTGGATTGAAGATGTAGATGGACTTGCAGAA 6975
Qy 2207 AAAATTATGAAGTTTCTAAGCAACTAATTTTACATCAAAATAAATTTTGTGCA 2266
Db 6976 GCGTTGAAAAGGAATATAGCTACAGAAAAATATCAGGGAATAATGATATGTTTGTGAT 7035
Qy 2267 AGATTAAAAACAATAGTTTCAAAAAAT 2292
Db 7036 AAAATTAGAAAAATATAGGTGAAT 7061
```

```
ID ADF43363 standard; DNA; 17276 BP.
XX AC ADF43363;
XX DT 12-FEB-2004 (first entry)
XX DE Sterptococcus capsular polysaccharide DNA seq id 83.
XX KW receptor; lipid-based tumour associated antigen; cytostatic;
XX KW antimicrobial; gene therapy; neoplastic disease; tumour; cancer;
XX KW infectious disease; capsular polysaccharide; ds.
XX OS Streptococcus.
XX PN US2003157113-A1.
XX PD 21-AUG-2003.
XX PF 28-DEC-2000; 2000US-007511708.
XX PR 28-DEC-1999; 99US-0173371P.
XX PA (TERM/) TERMAN D S.
XX PI Terman DS;
XX DR WPI; 2003-787326/74.
XX XX P-PSDB; ADF43364.
XX PT New receptor in a mammalian cell that inhibits regular activation by
XX PT receptors specific for lipid-based tumor associated antigens, useful for
XX PT treating a neoplastic disease or tumor, and infectious diseases.
XX PS Example 3; SEQ ID NO 83; 151pp; English.
XX CC The invention describes a receptor in a mammalian cell that inhibits
XX CC regular activation by receptors specific for lipid-based tumour
XX CC associated antigen. The receptor has cytostatic and antimicrobial
XX CC properties and is suitable for use in gene therapy. The receptors,
XX CC methods and compositions are useful for treating a neoplastic disease or
XX CC tumour (cancer), and infectious diseases. This sequence represents a
XX CC streptococcal capsular polysaccharide polynucleotide, a cell surface
XX CC moiety, the DNA of which can be transfected into a cell with superantigen
XX CC DNA to generate antitumour immunity.
SQ Sequence 17276 BP; 5953 A; 2414 C; 3043 G; 5866 T; 0 U; 0 Other;

Query Match 10.4%; Score 726.8; DB 10; Length 17276;
Best Local Similarity 59.2%; Pred. No. 9e-89;
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

Qy 16 TGGCATTTATTTGATGATAGCAGTTGCAATTTCTGCAATCTTAAACAGTCATATACCAA 75
Db 4765 TAGCGATTTCAACACAGTTGATGTTATTTTCTGCAAGTTTGACATTAACATTAATTA 4824
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATAATGATGTTTCATTATTTTGCAT 132
Db 4825 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATATATTTT 4884
Qy 133 TTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAAA 192
Db 4885 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 4944
Qy 193 AAAATTTAATATATATATATTTTGCATTTTCTTACGGCAGTATCATTTTGTGG 252
Db 4945 TGGTATTTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTTT 5004
Qy 253 AGAATAATTTCCGCACTTTCAAGACGTGGTCCGCTGTTATTTCCATTAATAAATCTCGTTT 312
Db 5005 AAACTCTTTTACAAACGACGACTTTCCCTTTTACTTTTACTTTTATGCTATGATTCGATTT 5064
Qy 313 TGGTATACCTATTTAAACGTAATTTAAGCAGTTTAAAGGATAGCTTTCTATTTTTCGACAA 372
```


Db 5065 TATTATATCTATTGAATTCATTTTTTAAATAATTATCGAABATATTTCTTACGCTAAGTTTT 5124
Qy 373 TCTATCAAAAAGACAGATCTTAATTACACGGCTGAACGATGGGAAAAATATGCAAGTTTT 432
Db 5125 CACGAGATACCAAAAGTTGTTTTGTATAACGATAGGATTTCTTTATCAAAAATGACCTTTA 5184
Qy 433 TATTGTAATCACATAAACAATAATCCAAAAATCTGTTGCAATGCGTAGTTTAAAGTACAG 492
Db 5185 GGAATAAAT---ACGACATAAATTATATCGCTGCTGTATCTTGGACTCTCTGAAAGG 5241
Qy 493 AAATAGATAAAATTAATTTATCATTTACCGCTCTATTATTCTGTGGAAGAAGCTATAGAGT 552
Db 5242 ATTGTTATGATTTGAAACATAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 5301
Qy 553 TTTCACACAGGGAAGTGTGACACAGCTCTTTATAAATCTACCAAGTGAGTTTTTAGACG 612
Db 5302 AGTTAAACCTGTTAACTGTTGATCAAGCTTTTATTAACATACCCATTGAAATTTATGGTA 5361
Qy 613 TAAAGCAA-----TTCGTTTCAGATTTTGAGTTGTTAGGTATTTGATGTAAGCGTTG 663
Db 5362 AATACCAATAACAAGATATTAATGACATTTGAAGCAATGGGAGTGATTTGTCATGTTA 5421
Qy 664 ATATTAAATTCATTCGGTTTACTCGTTTGAATAAACAATAAATCCAACTGCTAGGTGACC 723
Db 5422 ATGTAGAGGCACTTAGCTTTGATATATATAGAGAAAGCGAATCCAACTTTTGAAGAT 5481
Qy 724 ATAGATTTGAATCTTTTCCAAATTTTATAGCCCTAGTCATATCATGATGAAACGAC 783
Db 5482 ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT 5541
Qy 784 TTTTGGATATCTCGAGCGGTAGTCGGTTAAATTTTGTGGTATAGTTTCTATTTTGT 843
Db 5542 TTTTGGATATCACGGGTGCTATTATAGTTTGTCTATATGTCATATGTGGCAATTTTTC 5601
Qy 844 TAGTTCCAAATTTATCGTAGATGGTGGACCGGCTATTTTGTCTCAGAAACGAGTTCGAC 903
Db 5602 TAGTTCCGCAATCAGAAAGATGGTGGACCGGCTATCTTTTCTCAAAATAGATAGGTC 5661
Qy 904 AGAATGGACGATATTTTACATCTACAAAGTTTCGATCGATGATATGTTGATGCTGAGGAC 963
Db 5662 GTAATGGTAGGATTTTATAGTTCTATAAATTCAGATCAATCGAGTAGTAGCAACAAA 5721
Qy 964 GCAAAAAGACTTGTCTAGCCAAACAGATGCAAGGTGGGTATGTTTTAAATGGAA 1023
Db 5722 TTAAGAAAGATTTATTTAGTTTACCAATCAAAATGACAGG-----CTAATGTTTAAAGTTAGA 5776
Qy 1024 AAACGATCTAGAAATTTACTCCAAATGGACATTTTCATACGCAAAAACAAGTTTACAGAGT 1083
Db 5777 AGATGATCCTAGAAATTTACTAAATAGGAAATTTATTTTCG-AAAAACAAGCATAGATGAT 5835
Qy 1084 TACCACAGTTTATAATGTTTTTAATTTGGCGATATGAGTCTAGTTGGTACACGTCACCTA 1143
Db 5836 TGCCTCAATTTCTATAATGTTTTTAAAGCGGATATGAGTTTAGCAGGAACACGCGCTCCCA 5895
Qy 1144 CAGTTGATGATTTGAAAAATATCTCTGTTCAAGAGACGATGAGTTTAAACGAG 1203
Db 5896 CAGTTGATGAAATGAAAAGTATAATTCACGCAAGACGACGCGCTTAGTTTAAAGCCAG 5955
Qy 1204 GGATTTACAGGTCCTGGCAGGTTAGTGTCTGTAGTAATATCACAGACTTCGACACAGTAG 1263
Db 5956 GAATCACCTGGTTTGGCAATATCTGGTAGAATATATTTACGTGATTTTGTATGAAATCG 6015
Qy 1264 TTCGTTTGGACTTAGCATACATTTGATTAATTTGGACTATCTGCTCAGATATTAATAATTTTAT 1323
Db 6016 TAAAGTTAGATGTTCAATATATCAATGAATGCTCTATTTTGGTCAGATATTAAGATTTATC 6075
Qy 1324 TAAAGACAGTGAAGTTGTTATTTGTTGAGAGGGAAGTAAAGTA----- 1372
Db 6076 TCCTAAACATAAAGGTAGTTTTTACTCGGGACAGGAGCTTAAGTAAAGGTTAAGGTTGAAG 6135
Qy 1373 -----TATGAAGTTTCTGTTGGTCTGTTCTTACAGGGGACATTTTGACTCACTGTTATTT 1426
Db 6136 GAATATAATGAATAATTTGTTCTGGTTGTTCAAGTGGTGTCTATCTAGCACACTTGAACCT 6195

Qy 1427 GTTAAACCGTTTTTGAAGGAAGAAACGTTTTTTGGGTAAACATTTGATTAAGAGGATGC 1486
Db 6196 TTTGAAACCCATTTTGGGAAAAAGAGATAGGTTTTTGGGTAACTTTTGTATTAAGAGATGC 6255
Qy 1487 AAGAAAGTCTTTTCAAGAAATGAAATAATGTATCCATGTTACTTTTCCAAACAATCGCAATCT 1546
Db 6256 TAGGAGTATTTCAAGAGAGAGATTTGTATATCATTTGCTTCTTTCCAAACAACCGTATGT 6315
Qy 1547 CATTAATTTAGTGAATAATACTCTTTTAGCTTTCAAAATTTTACGTGATGAGAAACAGA 1606
Db 6316 CAAAAACTTGGTAAAAATACTATTCTAGCTTTTAAAGTCTTTAGAAAAAGAAAGACAGA 6375
Qy 1607 TGTATATTTTCACTGTTGGCGCGCTGCTGTCCTCTCTTTTACATCGAAAACTATT 1666
Db 6376 TGTATATCATATCTGTTGGCGCTGTGACGATGACCAATTTCTTTTATATGTTGTAAGTTATT 6435
Qy 1667 TGGAGCAAGACGATTTATATGAAGTATTTGATCGAGTTTAAATACTCTACATTAACCTGG 1726
Db 6436 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGATAGATAAACCAACTTTTGACAG 6495
Qy 1727 AAAAAGTATTTATCCGCTAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGT 1786
Db 6496 AAAATTTAGTGTATCTCTGTAACAGATAAATTTATTTGTTTCAGTGGGAAGAAATGAAGAAGT 6555
Qy 1787 ATATCTTAATCTATTAACTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAGT 1846
Db 6556 TTATCTCCTAGGCAATTAATTTAGGAGGAATTTTAAATGATTTTGTACAGTGGGACA 6615
Qy 1847 CATGAACAACAGTTTAAATCGATTTGATAAAGAGATTTGATTTTGAATAAATAATGGAAGT 1906
Db 6616 CATGAACAGAGTTCAACCGCTTTATTAAGAAGTTGATAGATTAAAGGACAGGTGCT 6675
Qy 1907 ATAACCGACGAAATATTTATTTCAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG 1966
Db 6676 ATTGATCAAGAAGTGTTCATTTCAACCGGTTTACTCAGACTTCGAACCTCAGAAATTTGCAG 6735
Qy 1967 TATAAAAAATTTCTCAGTTTCAAGAAGAAATGGAACAATATATTTAAACAATCAGAAGTAGTT 2026
Db 6736 TGGTCAAAATTTCTCTCATATGATGATGAACTCTTACATGAAAGAAGCTGAGATTTGTT 6795
Qy 2027 ATTTGGCAGGAGCGCGCTACTTTTATGAAATTCATTTATCCAAAGGAAAAACAATTA 2086
Db 6796 ATCACACATGGCGCCGACGAGCTTTATGTCAGTTATTTCTTTAGGAAATTTACCAGTT 6855
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGTTGAAACATGTAATGATCATCAAGTAGAGTTTGTGA 2146
Db 6856 GTTGTCTTAGGAGAAAGCAGTTTGGTGAACATATCAATGATCATCAATACAAATTTTGA 6915
Qy 2147 AGAAGAAATTTTACAGATAATAATATTTTATTTATAGAAAAATATAGATGATTTGTTGAA 2206
Db 6916 AAAAAAATTTGCCACCTGTATCCCTTGGCTTGGATTTGAAGATGTAGATGGAATTTGCCGAA 6975
Qy 2207 AAAATATTGAAGTTTCTAAGCAAACTAACTTTTACATCAAAATTAATAATTTTGTGAA 2266
Db 6976 GCGTTGAAAGGAATATAGCTACAGAAAAATATCAGGAAATATGATATGTTTGTGAT 7035
Qy 2267 AGATTAATAAATAAGTTGAAAAATTT 2292
Db 7036 AAATTAGAAAAAATATAGGTGAAT 7061

RESULT 9

ADO40235

ID ADO40235 standard; DNA; 25020 BP.

XX AC ADO40235;

XX XX

XX DT 15-JUL-2004 (first entry)

XX DE S. agalactiae capsular polysaccharide synthesis (cps) gene.

XX KW Group B Streptococci; GBS; capsular polysaccharide synthesis; cps;
DB


```

KW detection; ds; gene.
OS Streptococcus agalactiae.
XX
XX Key Location/Qualifiers
XX misc_binding 153..174
XX /tag= a
XX /bound_moiety= "S. agalactiae cps gene specific probe"
XX 546..581
XX /tag= b
XX /bound_moiety= "S. agalactiae cps gene specific probe"
XX 10100..10121
XX /tag= c
XX /bound_moiety= "S. agalactiae cps gene specific probe"
XX 10176..10198
XX /tag= d
XX /bound_moiety= "S. agalactiae cps gene specific probe"
XX 10219..10242
XX /tag= e
XX /bound_moiety= "S. agalactiae cps gene specific probe"
XX
XX US2004009574-A1.
XX
XX 15-JAN-2004.
XX
XX 09-JUL-2002; 2002US-00192280.
XX
XX 09-JUL-2002; 2002US-00192280.
XX
XX (DATT/) DATTAGUPTA N.
XX (SHAH/) SHAH K.
XX
XX Dattagupta N, Shah K;
XX
XX WPI: 2004-090471/09.
XX GENBANK; AFB028896.
XX
XX Novel oligonucleotide probes for detecting group B Streptococci e.g.
XX Streptococcus agalactiae in samples.
XX
XX Example 1; Fig 1; 28pp; English.
XX
XX The invention relates to compositions and methods for detecting group B
XX Streptococci (GBS) Streptococcus agalactiae capsular polysaccharide
XX synthesis (cps) gene in sample. The invention is useful for detecting S.
XX agalactiae in a sample. The present sequence is S. agalactiae cps gene.
XX This sequence is used to illustrate the method of the invention.
XX
XX SQ Sequence 25020 BP; 8592 A; 3630 C; 4594 G; 8204 T; 0 U; 0 Other;
XX
XX Query Match 10.3%; Score 721.4; DB 12; Length 25020;
XX Best Local Similarity 58.9%; Pred. No. 4.7e-88;
XX Matches 1389; Conservative 0; Mismatches 926; Indels 42; Gaps 7;
XX
XX 16 TGGCATTATTTGATATGATAGTAGCAGTTGCAATTTCTGCAATCTTAACAGAGTCATATACCAA 75
XX 5154 TAGCGATGATTCACACAGTTGGTGTATTTCTGCAAGTTTGACATTACATTAA 5213
XX
XX 76 ATGCTGATTTAAATCGTTCTGGAATTTT-----TATCATTAATGATGGTTTCATTATTGCA 131
XX 5214 CTCCTCAATTTTAAAGCAATAAAGATTATTTGTTGTTCTTAATTGATACATTATATGTC 5273
XX
XX 132 TTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGTAACTCATAGAGTTTCAA 191
XX 5274 TTTTATCTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAA 5333
XX
XX 192 AAAACATTAACTATAGTATAATATTTGCAATTTTCTTACGGCAGTATCATTTTGTG 251
XX 5334 ATGTTATTTGAATACAGCTTTTACTATATTTTCAATCAAGTTTCATTTTATTTT 5393
XX
XX 252 GAGAAATATTTGCACTTTTCAAGCGTGGTCCCGTGTATTTTCAATTAAATCACTCGTT 311
XX 5394 AAAAATCATTTACACGACGACTTTCCCTTTTTTTCTTTTATTTGCTATGAATTCGATT 5453

```

```

QY 312 TTGGTATACCTATTATTAACGTAAATTATTAAAGCAGTTTAAAGGATPAGCTTCTTATTTTCGACA 371
XX
XX DB 5454 TTATTGTATCTATTGAATTCATTTTAAAAATATTATCGAAAAATATTTCTTACGCTAAGTTT 5513
QY 372 ATCTATCAAAAAAGACGATTTCTAATTACAACGGCTGAACGATGGGAAAAATATGCAAGTT 431
XX
XX DB 5514 TCACGAGATACCAAGTTGTTTGTATTAACGAATGAAGATTTCTTTATCAAAAATACCTTT 5573
QY 432 TTATTGTGAATCACATAAAACAAATTCAAAAAATCTTGTGATTCGTTAGTTTGTAGGTAC 491
XX
XX DB 5574 AGGAATAAAT---ACGACCATAATATTATTGCTGCTGCACTTGGACTCCTCTGAAAG 5630
QY 492 GAAATAGATAAAATTAATTTATCAATACCGCTCTATTATTCTGTGGAAAGAGCTATAGAG 551
XX
XX DB 5631 GATTGTTATGATTGMAACATAAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCA 5690
QY 552 TTTTCAACMAGGAGTGGTCGACCACTCTTTATTAATCTACCAAGTGAAGTTTGTAGAC 611
XX
XX DB 5691 GAGTTAACTCTTAACTGTTGATCAAGCGTTTATTATAACATACCACTTGAATTTTGGT 5750
QY 612 GTAAAGCAATTCG-----TTTTCAGATTTTGAAGTTGTTAGGTATTGATGTAAGCGTT 662
XX
XX DB 5751 AAATACCAATACAAAGATATTATTAAATGACATTTGAACGATGGGAGTGAATGCTCAATGTT 5810
QY 663 GATATTAAATTCATTCGGTTTCTACTCGGTTGAAAAACAAAAAATCCAACTGCTAGGTGAC 722
XX
XX DB 5811 AATGTAGAGGCACCTTAGCTTTGATAATATAGGAGAAAGCGAATCCAACTTTTGAAGA 5870
QY 723 CATAGCATTTGTAACCTTTTCCACAAATTTTATTAAGCCTAGTCAATCATGATGAAGAACGA 782
XX
XX DB 5871 TATAGTGTATTACATATTCTATGAATTTCTATAAATATAGTACCTTTATAGCAAAACGA 5930
QY 783 CTTTTGGATATACCTCGGAGCGGTAGTCGGGTTAATTTATTTGTTGTTATAGTTTCTATTG 842
XX
XX DB 5931 TTTTGTGATTCACGGGTGCTATTATAGTTTGTCTCATATGTTGGCAATTTGGCAATTTT 5990
QY 843 TTAGTTCCAAATTTATTCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAACAGTTGGA 902
XX
XX DB 5991 CTAGTTCCAAATCAGAAAGATGGTGGACCGGCTATCTTTCTCAAAATAGAGTAGGT 6050
QY 903 CAGAAATGACGATTTTATACATTTCAAGTTTCATGATGATGATGTTGATGCTGAGGAG 962
XX
XX DB 6051 CGTAAATGTTAGGATTTTGTAGATTTCTATAAATTCAGATCAATCGGAGTAGATGACAGAACAA 6110
QY 963 CGCAAAAGAGACTTCTCAGCCAAACAGATCAAGGTTGGGTATGTTTAAATGGGA 1022
XX
XX DB 6111 ATTAAGAAAGATTTATTAGTTTCAATCAAAATGACAGGG-----CTAATGTTTAAAGTAG 6165
QY 1023 AAAACGATCCTAGAAATTTACTCCAATTTGACATTTTATACGCAAAAAACAAAGTTTACAGCAG 1082
XX
XX DB 6166 ACGATGATCCTAGAAATTTACTAAATAGGAAATTTATTTCG-AAAAACAGCATAGATGAG 6224
QY 1083 TTACCACAGTTTTTATAATTTTAAATTTGGCGATAGTCTAGTTGGTACAGTCCACCT 1142
XX
XX DB 6225 TTGCCTCAATTTCTATAATTTTAAAGTTGATATGAGTTTATAGTAGGAACACGCCCTCCC 6284
QY 1143 ACAGTTGATGAATTTGAAAAATATATCTCCTGGTCAAAAGACAGGATGAGTTTAAACCA 1202
XX
XX DB 6285 ACAGTTGATGAATATGAAAGTATAATTTCAACGAGAGCGACGCTTAGTTTAAAGCCA 6344
QY 1203 GGGATTAACAGTCTCTGCGCAGTTTAGTGGTCTAGTATATATACAGACTTCGACGACGTA 1262
XX
XX DB 6345 GGAATCACTGTTGTTGGCAATATCTGGTAGAATATATATCTGATTTTGTATGAATTC 6404
QY 1263 GTTCGGTTGGACTTTAGCATACATTTGATAATTTGGACTATCTGGTCAGATATTAAATTTTA 1322
XX
XX DB 6405 GTAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTTGGTCAGATATTAAAGTTAT 6464
QY 1323 TTAAGACAGTGAAAGTTGTATTGTTGAGAGGGAAGTAAAGTAAAGTA----- 1372
XX
XX DB 6465 CTCCTAACACTAAAGGAGTAGTCTTACTTTGGGACAGGTGCTAAAGTAAAGGTTTGAAG 6524

```


1373 -----TATGAGAGTTTGTGGTGGTCTTTCAGGGGAGACATTTGACCTCATTGTTATT 1425
 6525 GGAATATAAGTAAATTTGCTGGTTCAGTGGTGGTCACTAGACACTTGAACC 6584
 1426 TGTATAAACCGTTTGGGAGGAGGAAAGCACTTTTGGGTAAACATTTGATAAAGAGGATG 1485
 6585 TTTTGAACCACTTTGGGAGGAGGAAATAGATAGTTTGGGTAAACCTTTGATAAAGAGATG 6644
 1486 CAAGAAGTCTTTGGAAGATGAGGAAATAGTATCATGTTTCTTCCAAACAAATCGCAATC 1545
 6645 CTAGGAGTATTTCTAAGAGAGAGATTCGTATATCATCTGTTCTTCCAAACAAACCGTAATG 6704
 1546 TCATTAAATTTAGTGAAGATCTTCTTCTAGCTTTCAAAATTTTACGTGATGAGAAACGAG 1605
 6705 TCAAAACCTTGGTAAAGATCTATCTAGCTTTTAAAGGCTTTAGAAAGAGAACGACGAG 6764
 1606 ATGTTTATTTATTCATCTGGTGGCGCGTGTGCTGCCCTCTTTTACATCGGAAACCTAT 1665
 6765 ATGTTATCATATCATCTGGTGGCGCTGTAGCAGTACCATCTTTTATATTGTTAGTTAT 6824
 1666 TTGAGCAAGACGATTTATTTGAAGATTTTGAATCGAGTTAATAATCTACATTAATCTG 1725
 6825 TTGTTGTGAGACCGTTTATATAGAGGTTTTCGACAGATAGATAAAACCAACTTTGACAG 6884
 1726 GAAACTAGTTTATCCGTAACAGATATTTTATTGTTTTCAGTGGGAGAAATGAGAGG 1785
 6885 GAAATTTAGTATCTCTGTACAGATAAATTTATTTGTTTTCAGTGGGAGAAATGAGAAAG 6944
 1786 TATATCTTAATCTTATTAACCTTGGGAGTATTTTAAATGATTTTCTTAAACAGTAGGAAC 1845
 6945 TTTATCTTAGGCAATTAATTTAGGAGGATTTTAAATGATTTTCTTAAACAGTAGGAAC 7004
 1846 TCATGAACACAGTTTATTCGATTTGATAAAGAGATTTGATTTATTTGAAAAAATGGAAG 1905
 7005 ACATGAACAGCAGTTCAACCGTCTTATTAAGAGAGTTGATAGATTAAAGGAGCAGGTGC 7064
 1906 TATAACGACCAATATTTATTTCAACAGGATATCTGACTATATTTCCAGATATTTCCAA 1965
 7065 TATTGATCAAGAGTGTTCATTTCAACCGGTTACTCAGACTTTGAACCTCAGAAATTTGCA 7124
 1966 GTATAAAAAATTTCTCAGTTTCAAGAGAAATGGAACAATATATTAACAAATCAGAAAGTAGT 2025
 7125 GTGGTCAAAATTTCTCATATGATGATATAGAACTTTACATGAAAGAGCTGAGATTGT 7184
 2026 TATTGGCACGAGGCGCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAAACAAT 2085
 7185 TATCACACATGCGGTCGAGCAGCTTTATGAATGCACTTTCTAAGGAGGAAAAAACAAT 7244
 2086 ATTGTTCCCTAGACAAAGAGATGTTGACATGTAATGATCAATCAAGTAGAGTTGT 2145
 7245 TGTGGTTCCTAGACAAAGAGATTTGGAGAGCATGTGAATTAATCATCAGGTGGATTTTT 7304
 2146 AAGAAGAAATTTTACAAGATAATATTTTATTTATAGAAATATAGATGATTTGTTTGA 2205
 7305 GAAAGAGTTATCTTGAAATTTGAATTAGATATATTTTGAATATCAGTGAATAGAGAA 7364
 2206 AAAAATTTAGAGTTTCTTAAGCAAAAC---TAACTTTTACATCAAAATATATTTTGTG 2262
 7365 TATTATTAAAGGAAAAAATATATCTACTAGTAAAGTAAATATATCAAAAACATGATTTTG 7424
 2263 TGAAGATTAACCAATAGTTGAAATTTTAAATGAGGATCAAGAAATGATGAATTAATAA 2322
 7425 TTTCTCTTTCAAAATGAACTTTTCAAAACTATTTTGAATAAATATATTTTGTGGAGAA 7484
 2323 AAGATGCATATTTTGATA 2339
 7485 AAAAATTTGAATTAACA 7501

RESULT 10
 ADM79774
 ID ADM79774 standard; DNA; 2226 BP.
 XX

AC ADM79774;
 XX 03-JUN-2004 (first entry)
 XX Group B Streptococcus cpsD-cdsE-cpsF-cpsG partial consensus DNA sequence.
 DE group B streptococcus; GBS bacterium; cpsD; cpsE; cpsF; cpsG; cpsI gene;
 KW M gene; GBS infection; ds.
 XX Streptococcus sp. 'group B'.
 XX WO2003025216-A1.
 PN 27-MAR-2003.
 XX 18-SEP-2002; 2002WO-AU001281.
 PF 19-SEP-2001; 2001AU-00007749.
 XX (WSYD-) WESTERN SYDNEY AREA HEALTH SERVICE.
 PA Fanrong K, Gilbert G;
 XX WPI; 2003-381495/36.
 XX Typing a group B streptococcus (GBS) bacterium for diagnosing the GBS
 PT infections in pregnant women, elderly or immunocompromised patients by
 PT analyzing the sequence of the regions in the cpsD, cpsE, cpsF, cpsG or
 PT cpsI/M gene of the bacterium.
 XX Claim 2; Fig 1; 106pp; English.
 XX This invention relates to a novel method of typing a group B
 CC streptococcus (GBS) bacterium which comprises analysing the nucleotide
 CC sequence of one or more regions within the cpsD, cpsE, cpsF, cpsG and/or
 CC cpsI/M genes of the bacterium, where the regions comprise one or more
 CC nucleotides having sequences that vary between types. The method is
 CC useful for preparing a composition for serotyping and/or subtyping a GBS
 CC bacterium for diagnosing GBS infections in pregnant women, elderly and/or
 CC immunocompromised patients. The present sequence is that of a consensus
 CC DNA sequence for the 3' end of the cpsD-cpsE-cpsF and the 5' end of the
 CC cpsG sequences of group B Streptococci which is related to the method of
 CC the invention.
 XX SQ Sequence 2226 BP; 767 A; 293 C; 417 G; 749 T; 0 U; 0 Other;
 Query Match 9.3%; Score 649.2; DB 11; Length 2226;
 Best Local Similarity 59.5%; Pred. No. 2.5e-78;
 Matches 1223; Conservative 0; Mismatches 793; Indels 38; Gaps 6;
 QY 16 TGGCATTTATTTGATATGATAGCAGTTGTCGAATTTCTGCAATCTTAAACAAGTCATATACAA 75
 DB 182 TAGCGATGATTCAACACAGTTGTGGTTATTATTTCTGCAAGTTTCACATTAAACATTAATTA 241
 QY 76 ATCTGATTTAATATCG---TTCTGGAATTTTATCATATGATGTTTCATTATTTTCAT 132
 DB 242 CTCCTCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATATATTTGCT 301
 QY 133 TTTTATATCTCTGATGCCAGTTTGAATTTGAGTATAGAGTAAATCTGATAGAGTTTCAAA 192
 DB 302 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGCTATCTTGAAGAGTTTAAAA 361
 QY 193 AAAACATTTAACTAGTATATAATTTTGCATTTTCTTACGGCAGTATCATATTTTGTGG 252
 DB 362 TGGTATTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTCATTATTTTATTTTAA 421
 QY 253 AGAATAATTTTCGCACTTTCAAGAGGTTGTCGCGGTATTTTCAATTAATTAACCTTCCTT 312
 DB 422 AAAAATCTTTTACAAACGACAGCACTTTCTCTTTTACTTTTACTTTTATGCTATGATTCG 481
 QY 313 TGGTATACCTATTAACTGATTAATTAAGCAGTTTAAAGGATAGCTTTCTATTTTTCGACAA 372
 DB 482 TATTATATCTATTGAATTCATTTTTAAATATTTATCGAAAAATATTTCTACGCTAAGTTT 541


```
Qy 373 TCTATCAAAAAAGACGATCTTAATTAACAAGGCTGAACGATGGGAAATATGCAAGTTT 432
Db 542 CACGAGATACCAAGTCTGTTTGAATAACGAATAAGGATCTTTATCAAAAAATGACCTTTA 601
Qy 433 TATTGATACATATAAACAATTCAAAAAATCTTTGTTGCAATGGTAGTTTATAGTACAG 492
Db 602 GGAATAAAT---ACGACCATAATATATCGCTGTCTGTATCTTGACATCTCTGAAAAGG 658
Qy 493 AAATAGATAAAATTAATTTATCATTTACCGCTCTATTATTTCTGTGGAAGAGCTATAGAGT 552
Db 659 ATTGTTATGATTTGAAACATAACTCGTTTAAGGATAATAAACAAGATGCTCTTACTTCAG 718
Qy 553 TTTCAACAAGGGAAGTGTGCAACACGCTCTTTATAAATCTTACCAAGTGAGTTTTFAG--- 609
Db 719 AGTTAACTGCTTAACCTGTGATCAAGCTTTATTAACATACCCATCGAATTAATTTGGTA 778
Qy 610 -----ACGTAAGCAATTCGTTTCAGATTTTGTGTTAGTATTGATGTAAGCGTTG 663
Db 779 AATACCAATAACAAGATATTATTAATGACATTTGAAGCAATGGGAGTGATTTGTCATGTTA 838
Qy 664 ATATTAAATTCATTCGGTTTAACTGGTTTGA AAAACAAAAAATCCAACTGTAGTGACC 723
Db 839 ATGTAGGCACTTAGCTTTGATTAATATAGAGAAAGCGAATCCAACTTTTGAAGAT 898
Qy 724 ATAGCATTTGAACCTTTTCCAAATTTTATAAGCCTTAGTCATATCATGATGAACGAC 783
Db 899 ATAGTGTATTACATATTCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT 958
Qy 784 TTTTGGATATACCTGGAGGGTAGTCGGGTTAAATTTATTTGGTATAGTTCTATTTTGT 843
Db 959 TTTTGGATATCACGGGTGCTATTATAGGTTTGTCTATATGTGGCAATTTTGTGCAATTTTC 1018
Qy 844 TAGTTTCCAAATTTATCGTAGAGATGTGGACCGGCTATTTTCTCAGAAAACGAGTTGGAC 903
Db 1019 TAGTTTCCGCAATTCAGAAAGATGTGGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC 1078
Qy 904 AGAATGGACGCAATTTTACATTTCTCAAGTTTCGATCGATGATGTTGATGCTCAGAGAGC 963
Db 1079 GTAATGGTAGGATTTTATAGTTCTATATAATTCAGATCAATCGAGTAGATCGAGAACAA 1138
Qy 964 GCAAAAAAGACTGTCTAGCCGCAAAACAGATGCAAGGGTGGGTATGTTTTAAATGGGAA 1023
Db 1139 TTAAGAAAGATTTATTAGTTCAATCAATCAATGACAGGG-----CTAATGTTTAAAGTTAGA 1193
Qy 1024 AAACGATCCTAGAAATTTACCAATTTGGACATTTCTATCGCAAAACAGTTTAGACGAGT 1083
Db 1194 CGATGATCCTAGAAATTTACTAAATAATAGGAAAAATTTATTCTG-AAAAACAAGCATAGATGAGT 1252
Qy 1084 TACCACAGTTTTATATGTTTAAATTTGGCGATATCAGTCTAGTTGGTACACGTCACACCTA 1143
Db 1253 TGCCTCAATTTCTAATGTTTAAAGGTGATATGAGTTTAGTAGGAACAAGCCCTCCCA 1312
Qy 1144 CAGTTGATGAATTTGAAAAATATATCTCCTGGTCAAAAGAGAGCATTTAGTTTAAACAG 1203
Db 1313 CAGTTGATGAATATCAAAAGTATATAATCAACGCAAGCGAGCGCTTAGTTTAAAGCCAG 1372
Qy 1204 GGAATACAGGTTCTGGCAGGTTAGTGTGCTAGTAAATATACAGACTTCGACGACGTAG 1263
Db 1373 GAATCACTGGTTTGTGGCAATATCTGGTAGAAATTAATTAATCTGATTTTGTATGAATCG 1432
Qy 1264 TTTGGTTCGACTTAGCATATGATTAATTTGACATCTCTGGTACAGATTAATAAATTTTAT 1323
Db 1433 TAAAGTTAGATGTTCAATATATCAATGAATGTTCTATTTTGGTACAGATTAAGATTTATC 1492
Qy 1324 TAAAGACAGTCAAAAGTTGTTTGTGAGAGGGGAAGTAAGTAAAGTAT----- 1373
Db 1493 TCCTAACACTAAAGGTAGTCTTACTTTGGACAGGAGCTAAGTAAAGGTAAGGTTTGAAG 1552
Qy 1374 -----ATGAAAGTTTGTGTCGGTTCTTCAGGGGCAATTTGACTCAGTTGATTT 1426
Db 1553 GAATATAATGAAAAATTTCTGTTGTTTCAAGTGGTGTCTATCTAGCACACTTGAACCT 1612
```

```
Qy 1427 GTTAAACCCGTTTGGGAAGGAAGAACGTTTTTTGGGTAAACATTTGATAAAGAGGATGC 1486
Db 1613 TTTGAAACCCATTTGGGAAAAAGAGATAGGTTTTGGGTAACTTTTGATAAAGAGATGC 1672
Qy 1487 AAGAGTCTTTTGAAGATGA AAAAATCTATCCATGTTACTTTCCCAAAAATCCCAATCT 1546
Db 1673 TAGGAGTATTTCTAAGAGAGAGATTGTATATCATTTGCTCTTTCCAAACCCGTAATGT 1732
Qy 1547 CATTAATTTAGTGA AAAAATACTTTCTTAGCTTTTCAAAAATTTTACGTGATGAGAACCCAGA 1606
Db 1733 CAAAACTGTTGTA AAAAATACTATTCTTAGCTTTTAAAGTCTTAGAAAAAGAAAGACCCAGA 1792
Qy 1607 TGTATTATTTCATCTCGTGGCGCGTGTCTGTCCTCTTTTACATCGGAAACATATT 1666
Db 1793 TGTATTATCATCTCTCGTGGCGCTGTAGCAGTACCATTTCTTTATATTTGTAAGTTATT 1852
Qy 1667 TGGAGCAAGACGATTTATATTGAAGTATTTTGATCGAGTTAATAAATCTACATTAACCTGG 1726
Db 1853 TGGTTGTAAGACCGGTTTATATAGAGGTTTTCCGACAGATAGATAAACCACTTTTGACAGG 1912
Qy 1727 AAAACTAGTTTATCCCGTAAACAGATATTTTATTGTTTCAAGTGGGAAGAAAATCAAGAGGT 1786
Db 1913 AAAATTAGTGTATCTCTGAACAGATAAATTTATTGTTTCAAGTGGGAAGAAATGAAAAAGT 1972
Qy 1787 ATATCCTAAATCTAATAACTTTGGGAGTATTTTTTTAATGATTTTTTGAACAGTAGGAAC 1846
Db 1973 TTATCCTAAGGCAATTAATTTAGGAGGAAATTTTAAATGATTTTGTTCACAGTGGGACA 2032
Qy 1847 CATGAACAACAGTTTATCGATTGATTAAGAGATTTGATTTATTGCAAAAAAATCGAAGT 1906
Db 2033 CATGAACAGCAGTTTCAACCGCTTTATTAAAGAGTTGATAGATTAAAAAGGACAGGTGCT 2092
Qy 1907 ATAACCCAGCAAAATATTATTCAAAACAGGATATTCTGACTATATTCCAGAAATATTGCAAG 1966
Db 2093 ATTGATCAAGAGTGTTCATTCAACCGGTTACTCAGACTTTGAACCTCAGATTGTCTAG 2152
Qy 1967 TATAAAAAATTTTCAGTTTACAAAGAAATGGAACAATATATATAAACAATCAGAAAGTATT 2026
Db 2153 TGGTCAAAATTTCTCATATGATGATGATGAATCTTACATGAAAGAGCTGAGATTGTT 2212
Qy 2027 ATTTGCCACGAGG 2040
Db 2213 ATCACATGCGCG 2226

RESULT 11
AAH43197
ID AAH43197 standard; DNA; 6850 BP.
XX
AC AAH43197;
XX
DT 17-OCT-2001 (first entry)
XX
DE Portion of EPS352 plasmid.
XX
KW EPS plasmid; Lactococcus lactis subspecies cremoris Ropy352;
KW exopolysaccharide; EPS352; milk; thickener; glycosyltransferase;
KW biosynthesis; fruit juice; ds.
XX
OS Lactococcus lactis.
XX
FH Key Location/Qualifiers
FT CDS 174..491
FT /*tag= a
FT /product= "EpsR"
FT CDS 528..977
FT /*tag= b
FT /product= "EpsX"
FT /note= "No start codon given"
FT CDS 1020..1799
FT /*tag= c
FT /product= "EpsA"
FT CDS 1809..2504
```



```

FT      CDS      /*tag= d
FT      /*product= "EpsB"
FT      2618..3310
FT      /*tag= e
FT      /*product= "EpsC"
FT      3332..4018
FT      /*tag= f
FT      /*product= "EpsD"
FT      4022..4471
FT      /*tag= g
FT      /*product= "EpsE"
FT      4974..5681
FT      /*tag= h
FT      /*product= "EpsO"
FT      5687..6781
FT      /*tag= i
FT      /*product= "EpsP"
XX      WO200157234-A2.
XX      09-AUG-2001.
XX      02-FEB-2001; 2001WO-US003404.
XX      02-FEB-2000; 2000US-0179888P.
XX      16-OCT-2000; 2000US-0241098P.
XX      (UYOR-) UNIV OREGON STATE.
XX      (USDA ) US DEPT OF AGRICULTURE.
XX      Trempey JE, Knoshaug EP, Sandine WE, Ahlgren JA, Dierksen KP;
XX      WPT; 2001-488889/53.
XX      P-PSDB; AAB47416, AAB47417, AAB47418, AAB47419, AAB47420, AAB47421,
XX      AAB47422, AAB47423, AAB47424.
XX      New bacterium useful in pharmaceutical formulations, food products and
XX      beauty cosmetics, comprises characteristics of Lactococcus lactis
XX      cremoris Ropy 352.
XX      Disclosure; Page 50-58; 73pp; English.
XX      This sequence represents a fragment of the EPS plasmid derived from L.
XX      lactis subspecies cremoris Ropy352. The EPS plasmid is about 32 kb in
XX      size and encodes at least 13 active genes. The enzymes encoded by these
XX      genes allow the bacteria to produce an exopolysaccharide, designated
XX      EPS352. When EPS352 is expressed in or added to milk, it imparts highly
XX      desirable sensory characteristics to the milk, including making the milk
XX      very thick, with a very smooth mouth-feel, and slightly sweet with an
XX      obvious chewable-bite. Open reading frames (ORF's) M and N show homology
XX      to glycosyltransferase involved in EPS352 biosynthesis. L. lactis
XX      cremoris Ropy352 is deposited with the USDA-ARS-NCAUR-NRRL as deposit
XX      accession number NRRL B-30229. EPS352 is useful for thickening a liquid
XX      selected from milk, a milk-based liquid, a whey-based liquid, a soy-based
XX      liquid, and a fruit-juice. It is also useful as an additive in
XX      pharmaceutical products, beauty care products and coating agents. It is
XX      indicated, schematically, in the specification that there is an
XX      additional ORF between those encoding EpsE and EpsO. This encodes EpsF
XX      (AAB47425). However, this ORF is not indicated on the actual DNA sequence
XX      SQ      Sequence 6850 BP; 2487 A; 965 C; 1206 G; 2192 T; 0 U; 0 Other;
XX      Query Match      6.1%; Score 428.4; DB 4; Length 6850;
XX      Best Local Similarity 58.9%; Pred. No. 9.6e-49;
XX      Matches 849; Conservative 0; Mismatches 581; Indels 12; Gaps 6;
XX      861 AGAGATGGTGCACGGCTATTTTCTCAGAACGAGTTGGACAGATGGACGATATTT 920
XX      3518 AAAGATCAGGGCCAAATGTTCTATAAACAACAAACGCTATGGTAAATAATGGTAAATTTT 3577
XX      921 ACATTCTACAAGTTTCGATCGATGATGTTGATGCTGAGGAGCGCAAAAAGAGCTTGCTC 980
XX      3578 TATATTTTGAATTTAGAACAAATGATCTTAAATCCGAGCA--GTATCTAGAACTTAATC 3635

```

```

QY      981 AGCCAAAACAGATGCAAGSGTGGGTATGTTTTAAATATGGAAGAAACAGCTCTAGAAATTA 1040
DB      3636 CAGATGTTAAAGCTGCTTACCATGCCAACGGCAATAGCTAGAAAACAGATCCACGGGTAA 3695
QY      1041 CTCCAATTGGACATTTTCATACGCAAAAACAAAGTTTATAGACAGATTACACAGATTGTTATAAT 1100
DB      3696 CGAAGATTGGCTCATTTTATAAGACGACACTCAATT-GATGAACCTGCCACAATTTATCAAT 3754
QY      1101 GTTTTAAATTGGCGATATGAGTCTAGTTGGTACAGCTCCACCTACAGTTGATGATTTGAA 1160
DB      3755 GTTCTTTAAAGGGGATATGTCATTTAGTTGGTCCAAGACCAATTTCTGCTTTTGAAGCGAAA 3814
QY      1161 AATATATACCTCTGCTCAAAGAGACGATTTGAGTTTAAACCCAGGATTTACAGTCTCTGG 1220
DB      3815 GAATATGGGAAACGCTCTGCTTACTCATCTGTCAAACCCAGGAATCTCTGGTTATTGG 3874
QY      1221 CAGGTTAGTGGTCTAGTAAATATACAGACTTCGACGAGTCTGCTGGTTGGACTTAGCA 1280
DB      3875 ACGACATGCTGCAAGTAAAGTTCTTTTCTCAACGAGCAGAT---TTAGAATCTAT 3931
QY      1281 TACATTGATAATTGGACTATCTGGTCAGATATTAATAATTTTATAAGACAGTGAAGTT 1340
DB      3932 TATCTCAGTACCATAGCACCAAAAATGATATCAAGCTTCTAGTACTCAAAATTTGTACAA 3991
QY      1341 GTATTCTTGAGAGAGGGAAGTAAAGTATATGAAAGTTTGTGGTCTGGTTCTTCA 1400
DB      3992 AGTATTAAACGATCGGACGCATATTAATAA---ATGAAAATAGCATTTAGTAGTTCAGC 4048
QY      1401 GGGGACATTTGACTCACTTGTATTTGTTAAACCCGTTTGGAAAGAAAGAACGTTTTT 1460
DB      4049 GGTGGCCATTTGACACACCTGTTATTTGTTAAAAAAGTTTGGGAAAACGAAGATAGATT 4108
QY      1461 TGGGTAACTTTGATAAAGAGGATGCAAGAGTCTTTTGAAGAATGAAAAAATGTATFCCA 1520
DB      4109 TGGGTCACTTTGATAAACAAGATGCAAAATCTATATTGAAGAAGAAAGATTTTATCCCT 4168
QY      1521 TGTACTTCTCCACAAATCGCATCTCAATTAATTTAGTGAATAATCTTCTTAGCTTTC 1580
DB      4169 TGTATTATATCCCAAAATAGAAATGTAAAAACACGATAAAAAATACCATTTCTTGCATTT 4228
QY      1581 AAAATTTTACGTGATGAGAAACAGATGTTATTATTTTCACTCTGTCGGCGCTTGTCTGTC 1640
DB      4229 AAAATCTTTAGAAAAGAAAACACGATTTGATTTATTTTCGAGTGTGCTGCGGTAGCGTT 4288
QY      1641 CCCTTTCTTTTACATCGGAAAACCTATTTTGGAGCAAGACGATTTATATTGAAGTATTGAT 1700
DB      4289 CCTTTTCTTTTGGTTAGTAAACTATTTCGGTGCAAGACAGTCTATATTGAATAATTTGAC 4348
QY      1701 CGAGTTAATAAATCTACATTTACTGGAATACTAGTTTATCCCTGCTACAGATATTTTATTT 1760
DB      4349 CGGATCGATAAACCAACCTTTAACAGAAAATTTAGTTTATCCAGTTACTGATAAGTTTATA 4408
QY      1761 GTTCAGTGGGAAGAAATGAAGAAGTATATCTTAAATCTATTAACTTTGGCGAGTATTTT 1820
DB      4409 GTTCAATGGGAAGATTAAAAAAGTTTACCCTTAAGCAATTAATTTAGAGGAAATTTTC 4468
QY      1821 TAATGATTTTGTAAACAGTAGGAACCTCATGAAACAAAGTTTAAATTCGATTCGATAAAGAGA 1880
DB      4469 TAATGATTTTGTAAACGGTTGGAACTCACGAAACAAACCAATTAATTAATCGACTCATTTCAAAAA 4528
QY      1881 TTGATTTTAAAAAATAATGGAAGTATTAACCGACGAATATTTTATTTCAACAGGATATT 1940
DB      4529 TTGATGAACCTTTGACGCGATGGTGAATCGAAGACGATGTTATTTATGCAAAATTTGGGTACT 4588
QY      1941 CTGACTATATTCCAGAAATATTGCAAGTATATAAAAAATTTCTCAGTTACAAAAGAAATGGAAC 2000
DB      4589 CAATTTATGACCTAAATATCTAAATGGGAAAAGTTTATTTGATATGAGACTATGAAA 4648
QY      2001 AATATATTAACAATCAGAAGTAGTTTATTTGCCACGAGGCCCGCTACTTTTATGAATT 2060
DB      4649 GATGTATGAATGAAGCGGACGATGATTTATTTACTCATGCGGACCATCTTACCTATATGCAAG 4708

```


Db 861 GATTTCGAAAGGATGGGGATCTGCTATTTTGTCTCAGACGGGTATAGGAAAAAATGGTCG 920
Qy 914 CATATTTTACATTTCTACAGTTTCGATCGATGATATGTTGATGCTCAGAGCGGCAAAAAAGA 973
Db 921 TCAGTTTCACTTTTATAGTTTCGCTCTATGTGTAGATGCCGAGCGGCAAAAAAGAGA 980
Qy 974 CTTCGCTCAGCCAAACACAGATGCAAGGTTGGGTATGTTTAAATGCGAAAAACGATCCT 1033
Db 981 ACTCATGGAACAATAACCATGCAAGGTGGAAATGTTT-----AAGGTGGAGATGATCCT 1035
Qy 1034 AGAATTACTCCAATTTGGCAATTTTCATACGCAAAAAACAAGTTTAGACGAGTTTACACAGTT 1093
Db 1036 CGTATCAGAAAAATTGGTTGTTTATACG-GAAGACTAGCTTGCACGAGCTTACCACAGTT 1094
Qy 1094 TTATAATGTTTAAATTTGGCGATATGATCTAGTTGGTACAGTCCACCTACAGTTGATGA 1153
Db 1095 TTATAATGTTTAAAGGGAGATATGAGTTTGGTGGCAGCAGCGCCACCAACAGTGGACGA 1154
Qy 1154 ATTTGAAAAATATATCTCTCGTCAAAAGAGACGATTTGAGTTTAAACACGAGGATTACAGG 1213
Db 1155 GTATGAGCACTATACCCAGAACAAAAACGCCGACTAAGTTTAACTGCGATAACAGG 1214
Qy 1214 TCTCTGGCAGGTTAGTGTGCTAGTAAATATACAGACTTTCGACGACGTAGTTTCGGTTTGA 1273
Db 1215 TTTATGGCAGCTCAGCGGACGAAGTGAGATCAAGAATTTTCGATGAAGTTGTCAAATTAGA 1274
Qy 1274 CTTAGCATATGATTAATTTGGACTATCTGTCAGATATTAATAATTTTAAAGACAGT 1333
Db 1275 TGTGGCCCTATATTGATGGTTGGCAATCTGGAAGAATATTGAATTTTATTGGAAGACAGT 1334
Qy 1334 GAAAGTTGTATTGTTTGAGAGAGGGAAGTAAGTAA 1367
Db 1335 TAAAGTAGTATTTATGAGAGATGGAGCGAAGTAA 1368

RESULT 13

ID ADR93668 standard; DNA; 1443 BP.
XX ADR93668;

DT 16-DEC-2004 (first entry)

DE Novel S. pneumoniae DNA sequence, SEQ ID 2303.

XX Meningitis; bacteraemia; pneumonia; otitis media; ds;
KW bacterial infection.

XX Streptococcus pneumoniae.

XX US6800744-B1.

XX 05-OCT-2004.

XX 30-JUN-1998; 98US-00107433.

XX 02-JUL-1997; 97US-0051553P.

PR 12-MAY-1998; 98US-0085131P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2004-697205/68.

DR P-PSDB; ADR96271.

XX New isolated nucleic acid encoding a Streptococcus pneumoniae
PT polypeptide, useful for diagnosing, preventing and/or treating
PT pathological conditions resulting from the bacterial infection.

XX Disclosure; SEQ ID NO 2303; 151pp; English.

XX The invention relates to an isolated nucleic acid comprising a sequence

CC encoding a Streptococcus pneumoniae ADR91366peptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR9489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridisable under high stringency conditions to the nucleotide sequence.
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory
CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae nucleic acid sequences. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.

XX SQ Sequence 1443 BP; 432 A; 218 C; 340 G; 453 T; 0 U; 0 Other;

Query Match 5.9%; Score 415.8; DB 13; Length 1443;

Best Local Similarity 58.3%; Pred. No. 5e-47;
Matches 789; Conservative 0; Mismatches 552; Indels 12; Gaps 3;

Qy 14 ATTGGCAATATTGATGATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACC 73

Db 102 ATTGGCCATAGTCCAGAGTTTCTGTTATTTTATGTTGCTTACTTACTAGCAGTGTGAG 161

Qy 74 AAATGCTGATTTAAATCGTTCTGGAATTTTATCATTAATGATGGTTCATATTTTTCGATT 133

Db 162 AGAAACAGAGATTGTTTCAACAACAGCTATTGTACTTATATATCTCCACTATTTTGCCTT 221

Qy 134 TTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGTAACTGATAGAGTTTGAATA 193

Db 222 TTATATACAGTATTTGGAACAGGATTTCTTTAAAGGGGATATTTGATGAACCTTGCCA 281

Qy 194 AACATTTAACTATAGTATATATTTTGCATTTTCTTACGGCAGTATCATTTTGTGTGA 253

Db 282 GACATTTGAATATATCTTATCTTGGCTAGCGATAGTATTTCTTAATTTTCTTTAGA 341

Qy 254 GAATAATTTGCGCACTTTCAAGACGTGGTGGCGGTATTTTCCATTAATAAATCTTGTGTTT 313

Db 342 AGATCGATTTAGTATTTTCCAGAGAGCGCATGATTTACTTCTCTCTATTTACATGTTCTCTT 401

Qy 314 GGTATACCTATTTTAACGTAATTTAAGCAGTTTAAAGGATAGCTTCTATTTTCGACAAT 373

Db 402 AGTCTATGTCTAAACCGATTTTATCAAGTGGTATTTGGAACGGGCTTATCCCAACTTTAA 461

Qy 374 CTATCAAAAAAGACGATTTCTAATTAACAACGGCTGAACGATGGGAAATATGCAAGTTT 433

Db 462 AGGAAGTAAGAAGATTTCTCTACTTACAGCAACTTCTCGTGTGCAAAAGGATTTGGATAG 521

Qy 434 ATTTGAATCACTAATAAACAATTTCAAAAAATCTTTGTCATTTGTTAGTTTGTAGTACAGA 493

Db 522 ACTAATAGAATCAGATGATGTTCTGTTGGGAGTTGGTAGCGGTGAGTGTATTTAGATAAAC 581

Qy 494 AATAGATAAAATTAATTTATCATTCACGCTCTATTTATTTCTGTGGAAGAAGCTATAGATT 553

Db 582 AGATTTTTCAGCATGATTTATTTAAAGGTTGTAG-----CAGAGGGGGAGATCGTAAACTT 635

Qy 554 TTCAACAAAGGGAAGTGGTGCACCAACGCTTTTATAAATCTTACCAAGTGAGTTTGTAGACGT 613

Db 636 TGGACTCATGAGTGGTGCATGAAGTCTTTTATCAATCTTCCAGTGAAAAATACATAT 695

Qy 614 AAAGCAATTCGTTTTCAGATTTTTCAGTTGTTGATTTGATGTTAGCGGTTCATATTAATTC 673

Db 696 TGGAGAGCTTGTCTCTCAGTTTGAACGATGGGAGTTGATGTAAACAGTCAATCTAAATGC 755


```
QY 733 TAACTTTTCCACAAATTTTATAGCCTAGTATATATCATGATCAAAAGCATTTTGGATA 792
Db 5305 TTATGTTTTCACGATTTTATAGATATGAGATATCTTTGGCGAAGCGTTCCTCGATA 5364
QY 793 TACTCGGAGCGGTAGTCGGGTTAATTTCTGTGATATAGTTTCTATTTTGTAGTTCCAA 852
Db 5365 TCTGGGAGCTCTAGTTGTTGTTCTGTGATATGAGGAACTCTCTTTATCCTC 5424
QY 853 TTATTCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAACGAGTTGGACAGAAATCGAC 912
Db 5425 TTATTCGTAAAGATGGAGGACGAGTATTTTGTCTCAAGACCGTGTGGTGAATGGTC 5484
QY 913 GCATATTACATTTCTACAGTTTCGATCGATGATGTTGATCTGAGAGCGGCAAAAAG 972
Db 5485 GTATCTTTAAGTTTATAAATTCGTTCTATGCGTGTGATGCTGAGGAAATTAAGAACG 5544
QY 973 ACTTGCTCAGCAAAACAGATGCAAGGGTGGGTATGTTTAAATGGGAAAACGATCC 1032
Db 5545 AATTGATGGATAAAATCAANTGCTGGTGTATGTTT-----AGATAGACAATGATCC 5599
QY 1033 TAGAATTACTCCAATTGACATTTTCATACGCAAAAACAAGTTTAGACGAGTTACCAAGT 1092
Db 5600 ACGTATTACCAAGATTGGACATTTTCATTGCG-TAAAAACAAGTCTTGATGAACCTCCACAAT 5658
QY 1093 TTTTATAAGTTTAAATTTGGCGATATGAGTCTAGTTGGTACACGTCACCTACAGTTGATG 1152
Db 5659 TTTGGAATGTTCTAAAAGGTGATATGAGCTTTGGTGGAAACACGTCACCAACATTTGGATG 5718
QY 1153 AATTGTAAAAATATACTCTCGTCAAAAGACGACGATGAGTTTAAACCCAGGGATACAG 1212
Db 5719 AGTACGAACTTATACCCGGAACAAAACGTCGCCTCAGCTTTAAACCGAGTATTAATCTG 5778
QY 1213 GTCTCTCGCAGGTTAGTGGTTCGTAGTAATATACAGACTTTTCGACGACGTAGTTGGTTGG 1272
Db 5779 GTCTTTGCGAAGTAAGCGGTCGAAGTGAATTAATCTGATTTTCGATGAAGTTGTAAAACTAG 5838
QY 1273 ACTTAGCATATGATTAATTTGGACTATCTGCTCAGATATTAATTTTATTAAGACAG 1332
Db 5839 ATGTTGCTTATATGACGCGATGGCAATCTGGCGCGATATCAAAATCTTATTGAAAACGA 5898
QY 1333 TGAAGTTGTATCTGTAGAGAGGGAAGTAAGTAAAGTATATCAAGTTTGT 1386
Db 5899 TTAAGTAGTAGTAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 5952

RESULT 15
ID AA230355 standard; DNA; 19966 BP.
XX
AC AA230355;
XX
DT 11-FEB-2000 (first entry)
XX
DE Nucleotide sequence of the eps operon of S. thermophilus Sfi39.
XX
KW eps operon; Streptococcus thermophilus Sfi39; enzyme; eps1; eps2; eps3;
KW eps4; eps5; eps6; eps7; eps8; eps9; eps10;
KW exopolysaccharide biosynthesis; EPS; intersugar bond; antitumour;
KW probiotic; foodstuff; organoleptic quality; flavour;
KW lactic acid bacteria; acidified milk product; yoghurt; cheese; ss.
XX
OS Streptococcus thermophilus.
XX
FH Key Location/Qualifiers
FT CDS 2703..3389
FT /*tag= a
FT /product= "eps1"
FT /note= "encodes AAY43767"
FT CDS 3390..4121
FT /*tag= b
FT /product= "eps2"
FT /note= "encodes AAY43768"
FT

4130..4822
/*tag= c
/product= "eps3"
/note= "encodes AAY43769"
4832..5530
/*tag= d
/product= "eps4"
/note= "encodes AAY43770"
5629..6996
/*tag= e
/product= "eps5"
/note= "encodes AAY43771"
8633..10681
/*tag= f
/product= "eps6; this CDS, minus the termination codon,
is specifically claimed (claim 6)"
/note= "encodes AAY43772; contains a stop codon at
nucleotides 9410-9412"
10703..11671
/*tag= g
/product= "eps7"
/note= "encodes AAY43773; this CDS, minus the termination
codon, is specifically claimed (claim 6)"
12838..13788
/*tag= h
/product= "eps8"
/note= "encodes AAY43774; this CDS, minus the termination
codon, is specifically claimed (claim 6)"
14138..15553
/*tag= i
/product= "eps9"
/note= "encodes AAY43775; this CDS is specifically
claimed (claim 6)"
16919..18016
/*tag= j
/product= "eps10"
/note= "encodes AAY43776; this CDS, minus the termination
codon, is specifically claimed (claim 6)"
WO9954475-A2.
28-OCT-1999.
99WO-EP003011.
22-APR-1999; 98EP-00201310.
22-APR-1998; 98EP-00201311.
22-APR-1998; 98EP-00201312.
(NEST ) SOC PROD NESTLE SA.
Stingele F, Germond JE, Lamothe G;
WPI: 2000-013255/01.
P-PSDB; AAY43767, AAY43768, AAY43769, AAY43770, AAY43771, AAY43772,
AAY43773, AAY43774, AAY43775, AAY43776.
New recombinant enzymes for biosynthesis of exopolysaccharides having
e.g. antitumor or probiotic properties or useful in fermented milk
products.
Claim 6; Page 77-97; 163pp; French.
The present sequence represents the eps operon of Streptococcus
thermophilus Sfi39. The operon contains 10 open reading frames, and
encodes enzymes eps1, eps2, eps3, eps4, eps5, eps6, eps7, eps8, eps9 and
eps10 that are involved in the biosynthesis of exopolysaccharides (EPS).
The enzymes catalyse the formation of specific intersugar bonds. The
enzymes catalyse a process that includes at least one step of forming a
bond (in alpha or beta anomeric form) between C1, carrying the reducing
aldehyde group of an activated D-Galp (galactose in pyranose form), and a
phosphate group on a lipophilic or proteinaceous primer. The enzymes are
used to produce EPS that have antitumor or probiotic properties or are
```


CC used in foodstuffs to improve organoleptic qualities and flavour. When
CC expressed by lactic acid bacteria, EPS impart a free-flowing character
CC and/or a smooth, creamy texture to acidified milk products (yoghurt or
CC cheese)
XX
SQ Sequence 19966 BP; 6405 A; 3133 C; 3732 G; 6695 T; 0 U; 1 Other;
Query Match 5.4%; Score 376.2; DB 3; Length 19966;
Best Local Similarity 56.1%; Pred. No. 9.2e-42;
Matches 780; Conservative 0; Mismatches 593; Indels 18; Gaps 3;
2 TCGCAACAGAAATGGCAATATTGATATGATAGCAGTTGCAATTTCTGCAATCTTAAC 61
5649 TCGTCGCAATGNAATGGTATTATACAGTTGGTGTGGTGGTTCGAGCAATGGTAGC 5708
62 AAGTCATATACCAATGCTGATTTAAATCGTTCTGGAAATTTTATFATAATGATGGTTCA 121
5709 TAGTAAATATACCATATACAGAGATTACCAAGGAGCAATGTCCTTTAGGTGCTGACA 5768
122 TTATTTTGCATTTTATATCTCGTAGTCCAGTTGAATTTGAGTATAGAGTAATCTGAT 181
5769 TGTAGTGTCTTACTATATACAGTAGTTATATGAAATCTTAAGTATAGAGGCTACTTGA 5828
182 AGAGTTTGAAATTAATTAAGTATATGATATGATATGCAATTTTCTTACGGCAGTATC 241
5829 TGAATCTAATGCAATGTCATATATGTTTCTATATTTGCTCTAATTTGTAACATTTCTCTC 5888
242 ATTTTGTGAGAAATTAATTTGCACTTTCAAGACGTGTCGCGTGTATTTTCAATTAAT 301
5889 GTTTTTCAGATGGAAGTTTTCATCTCAGTCGCGACTCTTTACGTACCCCTGAT 5948
302 AAATCTGTTTGGTATACCTATTATTAACGTAATTAATTAAGCAGTTTAAGGATAGCTTCT 361
5949 TTCAGGTGTTCTTATACGTTTACAAATACATGTTCTTAAGTATTTCCGCTCATCTATTTA 6008
362 ATTTTCGACATCTATCAAAAAGACAGATTTCTAATTAACAGCGCTGAACATGGGAAA 421
6009 TACAGCTCGTAAAAAGTAAACAGATATTTCTTGTGATTTCTGATCAAGCACGCTAGAAA 6068
422 TATGCAAGTTTATTTGAATCACATAACAAATTCAAAAAATCTTGTTCATTTGGTAGT 481
6069 T-----GTTTTGCTCTGATGAAGACATATGATGGTAGGATACAGCAGT 6116
482 TTTAGGTACAGAAATAGATAAAATTAATTTATCATACCGCTCTATTTATTTCTGTGAGA 541
6117 TTGTGCTGTGATATCTTATTTTACCGATCCATTTATCAAGAGTGTAAACCTGAAA 6176
542 AGCTATAGATTTTCAACAGGAGTGGTGCACACGCTTTTATAAATCTACCAAGTGA 601
6177 TTTGATTTGAATATGCGACACACTCAGTAGTAGACCAAGTTTGTGATTAATCTGCCAAGTGG 6236
602 GTTTTTCAGCTAAGCAATTCGTTTCAGATTTTGGATTTGTAGTATTTGATGTAAGCGT 661
6237 GCAGTATGAATTTGGGATATATGATACCTTTTGGATATGGAATTCAGATTTCTAT 6296
662 TGATATTAATTAATTCATTCGTTTTTACTGCGTTTGAATAAACAATAAATTCAGTGTGTA 721
6297 TAATTTGAATGCCCTTGAATTTATGAATCAAGTGAATAAAGCGTATTCACAAATTTGGTCC 6356
722 CCATAGCATTTGATCTTTTTCACAAATTTTATAAGCCTAGTCAATCATGATGAACG 781
6357 TTTCAAAGTGTGTACGTTTTCACAGTATTTTATAGCTATGAGATATCTTTGGCGAAGC 6416
782 ACTTTTGTATATCTCGAGCGGTAGTGGGTTTATTTATTTTGTGTATAGTTTCTATTTT 841
6417 TTTCTCATATCTGTGGAGCTCTAGTTGGTTTGTGCTCTGTGGTATTTAGGAATCTT 6476
842 GTTAGTTCCTCAATTTATTCGTAGAGATGGTGGACCGCTATTTTGTCTCAGAAACGAGTTGG 901
6477 CCTTATCTCTTATTCGTAAAGATGGAGGACCGCAATTTTGTCTCAGACCGTGTGGG 6536
902 ACAGATGGACGCAATTTTACATTTCTACAAGTTTTCGATCGATGTATGTTGATGCTGAGGA 961

Db 6537 AGAAATATGGAGCTATCTTCAAGTTTATATAATTCGGTCTTATGTGTGTATGATCGGAAGA 6596
Qy 962 GCGCAAAAAGACATTTGCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTTAAATGGG 1021
Db 6597 AATCAAGAAGAAATTTGATGGCACAGAAATCAAAATGTCGTGCTATGTTT-----AAGATG 6651
Qy 1022 AAAACAGATCTAGAAATTTACTCCAAATTTGCAATTTTATACAGCAAAAACAAAGTTTATAGCA 1081
Db 6652 GACAATGATCCACGATTTTACCACAAATTTGGACATTTTCATTCG-TAAACACGAGTCTTGATGA 6710
Qy 1082 GTTACCAAGTATTTTATATGTTTAAATTTGGCGATATAGTCTAGTTGGTACAGCTCCACC 1141
Db 6711 ACTTCCACAAATTTGGATGTTCTTAAAGGTGATATAGCTTTGGTGGACACAGCTCTCC 6770
Qy 1142 TACAGTTGATGAATTTGAAAAATATATCTCTGGTCAAAAGAGACGATTTGAGTTTAAACC 1201
Db 6771 AACAGTTGATGATGATGAATAATATACACCTGAACAGAAACGTCGTTTAAAGTTTAAACC 6830
Qy 1202 AGGATATACAGTCTCTGGCAGGTTAGTGGTCTGTAGTATATATACAGACTTCGACGAGT 1261
Db 6831 TGGTATCAGTCTCTTTGGCAAGTAAGCGGTGCAAGTGAATTAATCTGATTTTGATGAAGT 6890
Qy 1262 AGTTCCGTTGACCTTAGCATATTTGATAATTTGGACTATCTGTCAGATTTTAAATTTT 1321
Db 6891 TGTAAACTAGACGTTGCTTATTTGGACGATGACAAATCTGGCGTGATATCAAAATCTT 6950
Qy 1322 ATTAAGACAGTGAAGTTGTATTTTGGAGAGAGGAAAGTAAGTAAAGTATATGAAAGT 1381
Db 6951 ATTGAAACGATTAAGTAGTAGTAGTAATCAAGGATGGAGCAAGTCATGCTTTCACCATT 7010
Qy 1382 TTGTTTGGTGG 1392
Db 7011 TCTTTTAATGG 7021
RESULT 16
AAZ45258
ID AAZ45258 standard; DNA; 19966 BP.
XX AC AAZ45258;
XX DT 27-MAR-2000 (first entry)
XX DE DNA encoding enzymes involved in exopolysaccharide biosynthesis.
KW Exopolysaccharide; EPS; ESP enzyme; EPS1; EPS2; EPS3; EPS4; EPS5; EPS6;
KW EPS7; EPS8; EPS9; EPS10; Streptococcus thermophilus strain sf139;
KW activated D-galactose pyranose; saccharide; beta-glycosyltransferase;
KW transporter; food; fermented milk product; yoghurt; cheese;
KW flavour stability; organoleptic property; ss.
XX Streptococcus thermophilus.
XX FH Location/Qualifiers
FT CDS 2703..3389
FT FT /*tag= a
FT FT /product= "EPS1"
FT FT /note= "encodes AAY54065"
FT CDS 3390..4121
FT FT /*tag= b
FT FT /product= "EPS2"
FT FT /note= "encodes AAY54066"
FT CDS 4130..4822
FT FT /*tag= c
FT FT /product= "EPS3"
FT FT /note= "encodes AAY54067"
FT CDS 4832..5530
FT FT /*tag= d
FT FT /product= "EPS4"
FT FT /note= "encodes AAY54068"
FT CDS 5629..6996
FT FT /*tag= e
FT FT /product= "EPS5"


```
FT CDS /note= "encodes AAY54069"
FT 8633..10681
FT /*tag= f
FT /product= "EPS6"
FT /note= "encodes AAY54070; contains 1 stop codon at
FT nucleotides 9410-9412"
FT CDS /tag= g
FT 10703..11671
FT /product= "EPS7"
FT /note= "encodes AAY54071"
FT 12838..13788
FT /*tag= h
FT /product= "EPS8"
FT /note= "encodes AAY54072"
FT 14138..15553
FT /*tag= i
FT /product= "EPS9"
FT /note= "encodes AAY54073"
FT 16919..18016
FT /*tag= j
FT /product= "EPS10"
FT /note= "encodes AAY54074"
XX WO9962316-A2.
XX 09-DEC-1999.
XX 22-APR-1999; 99WO-EP002841.
XX 22-APR-1998; 98EP-00201310.
XX 22-APR-1998; 98EP-00201311.
XX 22-APR-1998; 98EP-00201312.
XX (NEST ) SOC PROD NESTLE SA.
XX Stingle F, Germond JE, Lamothe G;
XX WPI; 2000-097267/08.
XX P-PSDB; AAY54065, AAY54066, AAY54067, AAY54068, AAY54069, AAY54070,
XX AAY54071, AAY54072, AAY54073, AAY54074.
XX New recombinant enzymes for synthesis of exopolysaccharides, particularly
XX in lactic acid bacteria, for improving properties of fermented milk
XX products.
XX Claim 6; Page 77-96; 162pp; French.
XX
XX The present sequence encodes enzymes involved in the biosynthesis of
XX exopolysaccharides (EPS). These enzymes are designated EPS1-EPS10. and
XX are encoded by open reading frames eps1-eps10. The enzymes are isolated
XX from Streptococcus thermophilus strain Sfi39. The proteins are used in a
XX method for the synthesis of EPS, which includes at least one step of
XX forming a bond (alpha or beta-isomer) between C-1 (carrying the reducing
XX aldehyde function, of an activated D-galactose pyranose), and a phosphate
XX on a lipophilic or proteinaceous primer. Stepwise synthesis of EPS occurs
XX with, in each step, addition of a new sugar unit, through its hemi-acetyl
XX function, to an alcoholic hydroxyl of a second sugar unit, present at the
XX end of a chain of sugar residues bonded to the primer. EPS1 to EPS4 are
XX used to elongate the EPS chain, and to regulate EPS production. EPS5
XX creates new bonds between saccharides, EPS6 is used in the biosynthesis
XX of EPS, EPS7 and EPS8 are beta-glycosyltransferases, EPS9 is a
XX transporter of repetitive units, and EPS10 catalyses the conversion of a
XX pyranose form of a beta-D-galactose to the furanose form. The EPS enzyme
XX are used to improve properties of foods, particularly fermented milk
XX products such as yoghurt and cheese, e.g. their organoleptic properties
XX and flavour stability
XX
XX Sequence 19966 BP; 6405 A; 3134 C; 3732 G; 6694 T; 0 U; 1 Other;
XX
XX Query Match 5.4%; Score 376.2; DB 3; Length 19966;
XX Best Local Similarity 56.1%; Pred. No. 9.2e-42;
XX Matches 780; Conservative 0; Mismatches 593; Indels 18; Gaps 3;
```


Db 6711 ACTTCCACAATTTGGAAATGTTCTAAAGGTGATGAGCTTGGTGGGACACGTCCTCC 6770
Qy 1142 TACAGTTGATGAAATTTGAAAAATATCTCTGGTCAAAAGAGACGATTTGAGTTTAAACC 1201
Db 6771 AACAGTTGATGATGAGAAATATACACCTGAACGAAACGTCGTTTAAAGTTTAAACC 6830
Qy 1202 AGGATTTACAGGTTCTCGGACGAGTTAGTGTGCTAGTAAATATACAGACTTCGACGAGT 1261
Db 6831 TGGTATCACCTGGTCTTTGGCAAGTAAGCGGTGCAAGTGAATTTACTGATTTGATGAAGT 6890
Qy 1262 AGTTCGGTTGAGCTTACATACATTTGATTAATTTGAGAGAGGGAAGTAAGTAAAGTATATGAAGT 1321
Db 6891 TGTAAGTACAGCTTCTCTTTTGGACGGATGACCAATCTGGCGTGATATCAAAATCTT 6950
Qy 1322 ATTAAGACAGTGAAGTCTGATTTGTTGAGAGAGGGAAGTAAGTAAAGTATATGAAGT 1381
Db 6951 ATTGAAGACGATTAAGTAGTAGTAAAGTAGTAAAGTAGTAAAGTAGTAAAGT 7010
Qy 1382 TTGTTGGTGC 1392
Db 7011 TCTTTAATGG 7021

RESULT 17

ABA01447
ID ABA01447 standard; DNA; 28170 BP.
AC ABA01447;
XX
XX
DT 21-FEB-2002 (first entry)
XX Streptococcus thermophilus eps10 operon #2.
DE Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
KW ds.
XX
XX Streptococcus thermophilus.
XX W0200179500-A2.
XX 25-OCT-2001.
XX 18-APR-2001; 2001WO-FR001199.
XX 18-APR-2000; 2000FR-00004972.
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
PA (DANO-) CIE DANONE SA GERVAIS.
PA (RHOD) RHODIA CHIM.
XX Rallu F, Bésancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX WPI; 2002-017616/02.
DR
XX
XX New nucleic acid fragments containing exopolysaccharide operon, useful
PT e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.
PT
XX Claim 9; Page 109-117; 144pp; French.
PS
XX The present sequence is an eps operon from Streptococcus thermophilus.
CC Proteins encoded by the eps operon function in exopolysaccharide (EPS)
CC synthesis. The operon is useful for producing chimeric eps operons, for
CC optimising production of EPS in lactic acid bacteria. EPS impart texture,
CC mouth feel and rheological properties to fermented food products (e.g.
CC yoghurt). They function as thickeners, to provide free-flowing and creamy
CC texture, and may also have biological activities beneficial to health
XX
SQ Sequence 28170 BP; 8951 A; 4464 C; 5468 G; 9287 T; 0 U; 0 Other;
Query Match 5.4%; Score 376; DB 6; Length 28170;
Best Local Similarity 56.7%; Pred. No. 9.7e-42;
Matches 782; Conservative 0; Mismatches 580; Indels 18; Gaps 4;

Qy 13 AATTGGCAATTTTGTATATGATAGCAGTTGCMAATTTCTGCAATCTTAAACAAGTCATATAC 72
Db 4600 AAATTGGTATTAACAGTTGGTTGGTTGTTTTCAGCCATGGTAGCTAGTAAATAC 4659
Qy 73 CAATGCTGATTTAAATCTGTTCTGGAATTTTATCATATATGATGCTTATTTTCAT 132
Db 4660 CTTATACAGAGATTTACCAAGGAAGCATTTGCTCTTTTAGGTGCTGATACATGATGCTT 4719
Qy 133 TTTTATATCTCTGATGCCAGTTGCAATTTGAGTATAGAGGTAATCTCTGATAGAGTTTGA 192
Db 4720 TCTATATCAGTAGTTTATATGAAAATCTTAAAGTATAGAGGCTACTTGGATGAATCTG 4779
Qy 193 AACATTTTAACTATAGTATTAATTTTGCATTTTCTTACGGCAGTATCAATTTTGTG 252
Db 4780 CAACTGTCAAATATTTGTTTCATATTTGCTCTAAATTTGTAATCTTCTCACCCTTTTTCAG 4839
Qy 253 AGAATTAATTTTCGACACTTTCAAGACGTTGGTGGTGGTATTTTCAATTAATAAATCTTCGTTT 312
Db 4840 ATGGAAGTTTCTCTATCTCAGTCGCGGACTTTTATGTTATCTCTGATTTCAAGTGTTC 4899
Qy 313 TGGTATACCTTATTTAAAGTAAATTTTAAAGCAGTTTAAAGGATAGCTTCTATTTTCGACAA 372
Db 4900 TCTTATACGTTACAAATACGTTCTTAAAGTATTTT-----CCGCTCTCTATCTATACAC 4953
Qy 373 TCTATCAAAAAAGACGATTTCTAAATTAACAACGGCTGAACGATGGGAAAATATCAAGTTT 432
Db 4954 GTCGTAAGAACATAAGAAATATTTCTCTGATTTCTGATCAAGCAGCTCTAGAAAAATGTTT 5013
Qy 433 TATTGATCAATCAAAACAAATTTCAAAAAATCTCTGTTGATTTGGTAGTTTGTAGTACAG 492
Db 5014 TGTCTCGTATGAAGACAAATATGATGGTAGGATTAACAGCCGTTTGTGTCTGGATATC 5073
Qy 493 AATAGATAAAATTTAAATTTTATCAATACCGCTCTATTTATCTGTGGAAGAGCTATACAGT 552
Db 5074 C-----TTATTTTACCGATCCATTTTCAAGAGTGTAAACCTGAAAATTTGATGAT 5127
Qy 553 TTTCAACAGGGAAGTGGTGCACACGCTCTTTATAATCTTACCAAGTGGAGTTTGTAGACG 612
Db 5128 ATCGACACACTCAGTAGTAGACCAAGTTTGTATTAATCTTCCAAAGTGACGAGTACAAGA 5187
Qy 613 TAAAGCAATTCGTTTCAGATTTTGGATTTGATGTTAGTATGATGTAAGGTTGATTAATTT 672
Db 5188 TTTGGGATTTACGATCACTTTTGAAGTTATGGGGATTTCCAGTATCCATTTAAATTTGAATG 5247
Qy 673 CATTCGTTTACTGCTGTAAGAAACAAACAAATCCAACTGCTAGTGGACCATAGCATTG 732
Db 5248 CCTCGAATTTATGAGTAAGGTTGAAAACGATATCCAACTTTGGGTCCTTTCAAAGTTG 5307
Qy 733 TAACTTTTTCACAAATTTTATAAGCTTAGTCTATATCATGATGAACAGCTTTTGGATA 792
Db 5308 TTACGTTTTCACGCAATTTTATAGCTATGGAGATATCTTGGCGAAACGTTTCTCGATA 5367
Qy 793 TACTCGAGCGGTAGTCGGGTTAAATTTTGTGGTATAGTTTCTATTTTGTAGTTCCAA 852
Db 5368 TCTGTGGAGCCCTAGTTGGTTTGGTCTCTGTGGGATTTGTGAAATCTTCTTTATCCAC 5427
Qy 853 TTATTCGTAGAGATGTTGGACCGGCTATTTTCTCAGAACAGATTTGGACAGAAATGGAC 912
Db 5428 TTATTCGTAGGATGTTGGGCGACCAATTTTCTCAAGACCGTGGGGAAGAAATGGAC 5487
Qy 913 GCATATTTTACATTTCAAGTTTTCGATGATGATGTTGATGCTGAGGAGCGCAAAAAG 972
Db 5488 GTATCTTCAAGTTTATAATTTCCGTTCTATGTTGTTGATGCGGAGAAATCAAGAGA 5547
Qy 973 ACTTGCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTAAAAATGGGAAAAACGATCC 1032
Db 5548 ATTTGATGGCAGAGATCAAAATCTCTGGTGTATGTTT-----AAGATGCAATGATCC 5602
Qy 1033 TAGAATTTACTCCAATTTGGACATTTTATACGCAAAAACAGTTTGTAGACGAGTTACCACAGT 1092
Db 5603 ACGTATTTACCAAAATTTGGACATTTTCATTCG-TAAAAACGAGTCTTGATGAATCTTCCAAAT 5661

Db 5657 TTTGGAAATGTTCTAAAAGGTGATATGAGCTTTGGTAGGAACAGCGTCCACCAACATTTGGATG 5716
Qy 1153 AATTTGAAAATATATCTCTGGTCAAAAAGAGACGATTTGAGTTTAAACCGAGGATTTACAG 1212
Db 5717 AGTACGAATCTTATACACCGGAACAAAACGTCGCCTCAGCTTTTAAACCGAGGATTTACTG 5776
Qy 1213 GTCTCTCGCAGGTTAGTGGTCTAGTAATATATCAACAGACTTTCGAGCAGTGTAGTTTCGGTTGG 1272
Db 5777 GTCTTTGGCAAGTAAGCGGTGCAAGTGAATATCTGATTTTGTGAAAGTTGTAAACCTAG 5836
Qy 1273 ACTTAGCATACATTTGATAAATTTGGACTATCTGGTCAGATATTTAAATTTTATTAAGACAG 1332
Db 5837 ACGTTGCTTATTTGGACGGATGGACATCTGGCGGATATTAATCTTATTTGAACAA 5896
Qy 1333 TGAAGTTGTATTTTGGAGAGAGGGAAGTAGTAAGTAAGTAAGTAAGTAAGTAAGTTTGGTTCG 1392
Db 5897 TTAAGTAGTAGTAATGAAGGATGGAGCAAGTAGTGGCTTTCTCCATTTCTTTAATGG 5956

RESULT 20
ABA01444
ID ABA01444 standard; DNA; 18488 BP.
XX ABA01444;
XX
XX
DT 21-FEB-2002 (first entry)
XX
XX Streptococcus thermophilus eps6 operon #2.
XX
XX Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
XX ds.
XX
XX Streptococcus thermophilus.
XX
XX WO200179500-A2.
XX
XX 25-OCT-2001.
XX
XX 18-APR-2001; 2001WO-FR001199.
XX
XX 18-APR-2000; 2000FR-00004972.
XX
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX (DANO-) CIE DANONE SA GERVAIS.
XX (RHOD) RHODIA CHIM.
XX
XX Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX
XX WPI; 2002-017616/02.
XX
XX New nucleic acid fragments containing exopolysaccharide operon, useful
XX e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.
XX
XX Claim 9; Page 94-99; 144pp; French.
XX
XX The present sequence is an eps operon from Streptococcus thermophilus.
XX proteins encoded by the eps operon function in exopolysaccharide (EPS)
XX synthesis. The operon is useful for producing chimeric eps operons, for
XX optimising production of EPS in lactic acid bacteria. EPS impart texture,
XX mouth feel and rheological properties to fermented food products (e.g.
XX yoghurt). They function as thickeners, to provide free-flowing and creamy
XX texture, and may also have biological activities beneficial to health
XX
SQ Sequence 18488 BP; 5800 A; 2982 C; 3655 G; 6051 T; 0 U; 0 Other;

Query Match 5.3%; Score 371.4; DB 6; Length 18488;
Best Local Similarity 55.9%; Pred. No. 4.1e-41;
Matches 777; Conservative 0; Mismatches 596; Indels 18; Gaps 3;

Qy 2 TCGCCAAACGAAATGGCATTATTTGATATATGATAGCAGTTGCAATTTCTGCAATCTTAAAC 61
Db 4589 TCGTCGCATTGAAATTTGATATATACAGTTGGTTGGTTGTTTCGCAGCATTGGTAGC 4648

Qy 62 AAGTCATATACCAAAATGCTGATTTAAATCGTTCTCGAAATTTTATCATATATGATGGTTCA 121
Db 4649 TAGTAAATATACCATATACAGAGATTTACCAAGAGATTTTGTCTCTTTTAGGTGCTGACA 4708
Qy 122 TTATTTTGTCAATTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGAT 181
Db 4709 TGTAGTGTCTTACTATATCAGTAGTTTATTAAGAAAATATTAAGTATAGAGGCTACTTGA 4768
Qy 182 AGAGTTTGAAGAAAACATTTAACTATAGTATATATTTGCAATTTTCTTTACGGCAGTATC 241
Db 4769 TGAATTCATTTGCAACTGTCAAAATATTTTTCATATTTTGTCTTAATTTGCAACTTTTCTCTC 4828
Qy 242 ATTTTGTGGAGAAATTTTCGACTTTCAAGACGTGGTGGCGTGTATTTTCAATTAAT 301
Db 4829 GTTTTGTGAGATGGGAAGTTTTCATCTCACGTGCGGACTTCTTTACGTCAACCTGAT 4888
Qy 302 AAACCTTGTGTGTATACCTATTTAAACGTAAATTTAAAGCAGTTTAAAGGATAGCTTTCT 361
Db 4889 TTCAGGTGTCTCTTATACGTTACAATACTGTCTTAAGTATTTCCGTTCACTATTTA 4948
Qy 362 ATTTTCGACATCTATCAAAAAAGACGATTTCTAAATTAACAACGGCTGAACGATGGGAAA 421
Db 4949 TACACGTCGTAAGTAAGTAACAAGAAATTTCTCTTGTATTTCTGATCAACAGCAGCTCTAGAAA 5008
Qy 422 TATGCAAGTTTATTTGAATCACATAAACAATTTCAAAAAAATCTTGTTCATTTGGTAGT 481
Db 5009 TGTTTTGTCTTGTATGAAAGACAATATGATGTTAGGATTTA-----CAGCAGT 5056
Qy 482 TTTAGGTACAGAAATAGATAAAATTAATTTATCATTCACGCTCTATTTATTTCTGTGGAAGA 541
Db 5057 TTGTGTCTGGATAATCCCTTATTTTACCGATCCATTTATCAAGAGCTTAAACCTGAAA 5116
Qy 542 AGCTATAGAGTTTTCACAAAGGAAGTGTGACACAGCTTTTATATAAATCTACAAGTGA 601
Db 5117 TTTGATTGAATATGCGACACACTCAGTAGTAGACCAAGTTTGTATTAATTTGCGGAGTGA 5176
Qy 602 GTTTTGTAGAGTAAGCAATTCGTTTCAGATTTTGTAGTTTGTAGTATTTGATGTTAGCGT 661
Db 5177 GCAGTACAAGATTTGGGATTTATGCGTCACCATTTGAACTTTATGGGAATCCAGTATCCAT 5236
Qy 662 TGATATTAATTCATTCGTTTACTCGTTGTAAGAAAACAAAAAAATCCAACTGCTAGTGA 721
Db 5237 TAATTTGAATGCCCTTGAATTTATAGTCAGGGTGAAAACGATTCACAACTTTGGTCC 5296
Qy 722 CCATAGCATTTGTAATCTTTTCCAAATTTTATAAGCCCTAGTCATATATATGATGAACG 781
Db 5297 TTTCAAAAGTTGTACGTTTCTACGCAATTTTATAGCTATGAGATATCTTTGGCAAAACG 5356
Qy 782 ACTTTTGGATATACTCGAGCGGTAGTTCGGTTAAATTTATTTGTGGTATAGTTTCTATTTT 841
Db 5357 CTTTCTCGATATCTGTGGAGCTCTAGTTGGTTTGGTCTCTGGTATTTGTTGGAATCTT 5416
Qy 842 GTTAGTTTCCAAATTTATTCGTAGAGATGTTGACCGGCTATTTTGTCTCAGAAAACGAGTTGG 901
Db 5417 CCTTTATCTCTTATTCGTAAAGTGGAGACCAAGCTATTTTGTCTCAAGACCGTGTGGG 5476
Qy 902 ACAGAAATGGACGATATTTACATTTACAGTTTTCGATCGATGATATGTTGATGCTGAGGA 961
Db 5477 AGAAATGGACGATCTCTCAAGTTTATATAATTCGTTCTATGATGTTGATGCGGAAGA 5536
Qy 962 GCCAAAAAAGACTTGTCTAGCCAAACACAGATGCAAGGGTGGGTATGTTTAAATTTGGG 1021
Db 5537 AATCAAGAGAATTTGTATGTCACAGAAATCAAAATGTCTGTGGTATGTTT-----AAGATG 5591
Qy 1022 AAAAAACGATCTAGAAATTTACTCCAAATTTGGACATTTTCATACGCAAAAAACAAGTTTAGACGA 1081
Db 5592 GACATATGATCCAGTATTAACCAAAATTTGACATTTTCAATTCG-TAAACAGAGTCTTGATGA 5650
Qy 1082 GTTTACACAGTTTATTAATTTTAAATTTGGCGATATGAGTCTAGTTGGTATGATGCTCAACC 1141
Db 5651 ACTTCCACAATTTTGGAAATGTTCTTAAAGGTGATATGAGCTTGTGTAGGAACACGCTCCACC 5710
Qy 1142 TACAGTTGATGAATTTGAAAAAATATACTCTCTGTGTCAAAAAGAGACGATTTGATTTTAAACC 1201


```
Db 5711 AACATTGGATGAGTACGAATCTTATACACCGGAACAAAAACGTCGCCCTCAGCTTTTAAACCC 5770
Qy 1202 AGGGATTACAGGTCTCTGCGAGTTAGTGTCTAGTAATATACAGAGCTTTCGACGAGT 1261
Db 5771 AGGTATTACTGTCTTTGGCAAGTAAGCGGTGCAAGTGAATTAATCTGATTTTGATGAAGT 5830
Qy 1262 AGTTCGGTTGGACTTTCAGATACATTTGAAATTAATTTGGACTATCTGGTCAGATATTTAAATTTT 1321
Db 5831 TGTAAACTAGACATTTGCTTATATGACGAGTGGACAATCTGGCGTGATATCAAAATATT 5890
Qy 1322 ATTAAGACAGTGAAGTTGATTTGTTGAGAGAGGGAAGTAAGTAAAAAGTATATCAAGT 1381
Db 5891 ATTGAACAATTAAGTAGTAGTAATGAAGATGAGCAAAAGTATGCGCTTTCTCCATT 5950
Qy 1382 TTGTTTGGTCG 1392
Db 5951 TCTTTTAATGG 5961

RESULT 21
ABA01436
ID ABA01436 standard; DNA; 19738 BP.
XX AC ABA01436;
XX 21-FEB-2002 (first entry)
XX Streptococcus thermophilus eps6 operon #1.
XX Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
XX ds.
XX Streptococcus thermophilus.
XX WO200179500-A2.
XX 25-OCT-2001.
XX 18-APR-2001; 2001WO-PR001199.
XX 18-APR-2000; 2000FR-00004972.
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX (DANO-) CIE DANONE SA GERVAIS.
XX (RHOD ) RHODIA CHIM.
XX Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX WPI; 2002-017616/02.
XX New nucleic acid fragments containing exopolysaccharide operon, useful
XX e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.
XX Claim 11; Page 52-57; 144pp; French.
XX The present sequence is an eps operon from Streptococcus thermophilus.
XX Proteins encoded by the eps operon function in exopolysaccharide (EPS)
XX synthesis. The operon is useful for producing chimeric eps operons, for
XX optimising production of EPS in lactic acid bacteria. EPS impart texture,
XX mouth feel and rheological properties to fermented food products (e.g.
XX yoghurt). They function as thickeners, to provide free-flowing and creamy
XX texture, and may also have biological activities beneficial to health
XX Sequence 19738 BP; 6140 A; 3173 C; 3904 G; 6422 T; 0 U; 99 Other;
XX Query Match 5.3%; Score 371.4; DB 6; Length 19738;
XX Best Local Similarity 55.9%; Pred. No. 4.1e-41;
XX Matches 777; Conservative 0; Mismatches 596; Indels 18; Gaps 3;
Qy 2 TCGCAAAACGAATTTGGCATTTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAAC 61
Db 4563 TCGTCGCAATTTGAAATTTGGTATTATACAGTTGGTTGGTTGTTTGGCGCATGGTAGC 4622
```

```
Qy 62 AAGTCATATACCAATGCTGATTTTAAATCGTTCTGGAATTTTATCATTAATGATGGTTCA 121
Db 4623 TAGTAAATAACTCTTATACAGAGATTACCAAGGAAGCATTTGTCCTTTAGGTGTCGTACA 4682
Qy 122 TTATTTTGCATTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGGTAATCTGAT 181
Db 4683 TGTAGTGTCTTACTATATACAGTAGTTAATTGAAAAATATTAAAGTATAGAGGCTACTTGA 4742
Qy 182 AGAGTTTGAAGAAACATTTAACTATAGTATATATTTCGAATTTTCTTTCGCGCAGTATC 241
Db 4743 TGAATTCATTCCAATCTGCTCAATATTGTTTCATATTGCTCTAATTCGCACTTTCTCTC 4802
Qy 242 ATTTTGTGTGGAATTAATTTTCGCACTTTCAAGACGTCGTCGCGTGTATTTTCAATTAAT 301
Db 4803 GTTTTGTGCAGATGGAAGTTTCTTCTATCTCAGTCGCGGACTTTCTTACGTCACCCCTGAT 4862
Qy 302 AAACCTCGTTTGTGTATACCTATTTAACGTAAATTTAAGCAGTTTAAAGGATAGCTTTCT 361
Db 4863 TTCAGGTGTTCTCTTATACGTTTACAAATPACTGTTCTTAAAGTATTTCCGTTCATCTATT 4922
Qy 362 ATTTTCGACAACTATCAAAAAAGAGCGATTTCTAATTACAAACGCTGAACGATGGGAAA 421
Db 4923 TACACGTCGTAAGAGTAACAAGAAATATTTCTTGTGATTTCTGATCAAGCAGCTGTAGAAA 4982
Qy 422 TATGCAAGTTTATTTTGAATCACATAAAACAAATTCAAAAAATCTTGTTCATTCGTTAGT 481
Db 4983 TGTTTGTCCTTGTATGAAAGCAATATCGATGGTAGGATTA-----CAGCAGT 5030
Qy 482 TTTAGGTACAGAAATAGATATAAATTAATTTATCAATCCGCTCTATATTCTGTGGAGA 541
Db 5031 TTGTCTCTGGTAATCTCTTATTTTACCGATCCATTTTCAAGAGTGTAAAAACCTGAAAA 5090
Qy 542 AGCTATAGAGTTTCAACAAGGAGTGTGCGACACGCTTTTATAAATCTTACCAAGTGA 601
Db 5091 TTTGATTAATATGCGACACACTCAGTAGTAGACCAAGTTTGTATTAATCTGCCGAGTGA 5150
Qy 602 GTTTTACACGTAAGCAATTCGTTTCAGATTTTTCAGTTGTGTAGTATGTAGTGAAGCT 661
Db 5151 GCAGTACAAGATTTGGGATTTATGCGTCACCATTTGAACCTTATGGGAATCCAGTATCCAT 5210
Qy 662 TGATTAATTAATTCATTCGGTTTTTACTCGGTTGAAAAACAAAAATCCAACTGTAGTGA 721
Db 5211 TAAATTTGAATGCCCTTGAATTTTATGAGTCAGGCGTAAAAACGATATCCAACTGCGTCC 5270
Qy 722 CCATAGCAATTTGTAATTTTCCACAAATTTTATAAGCCTAGTCATATCATGATCAAAACG 781
Db 5271 TTTCAAGTTGTTAGCTTTTCTACGCAATTTTATAGCTATGAGATATCTTTGGCGAAACG 5330
Qy 782 ACTTTTGGATATCTCGGAGCGGTAGTCGGGTTTAATTTTGTGTATAGTTTCTATTATT 841
Db 5331 CTTCTCGATATCTGTGGAGCTCTAGTTGGTTGGTGTCTGTGGTATTTTGGTAATCTT 5390
Qy 842 GTTAGTTTCCAAATTTTCGTAGAGATGGTGGACCGCTATTTTGTCTCAGAAACGAGTTGG 901
Db 5391 CCTTATCTCTTATTCGTAAAGATGGAGGACCAAGCTATTTTGTCTCAAGACCGTGTGG 5450
Qy 902 ACAGATGGACGATATTTTACATTTCAAGTTTTCGATCGATGTATGTATGTCTGTAGGA 961
Db 5451 AGAAAAATGGAGTATCTTCAAGTTTATAAATTCGTTCTATGCAATGTGTATGTGCGGAA 5510
Qy 962 GCGCAAAAAACACTGCTCAGCCAAACCAAGTCAAGGTGGGTATGTTTAAATGGG 1021
Db 5511 AATCAAGAAAGAAATTTGATGGCACAAGATCAAAATGTCTGTGTATGTTT-----AAGATG 5565
Qy 1022 AAAAACGATCTAGAAATTTACTCCAAATTTGCAATTTTATACGCAAAAAACAAGTTAGACGA 1081
Db 5566 GACATGATCACGATTTATCCAAATTTGACATTTTCATTCG-TAAAAACGAGTCTGTATGA 5624
Qy 1082 GTTACCACAGTTTATTAATGTTTTTAATTTGGCGATATAGTCTAGTTGTGTACACGTTCCAC 1141
Db 5625 ACTTCCACAAATTTGGTAATGTTTCTAAAAGGTGATAGTACTGTTGGTAGGAAACACGTTCCACC 5684
```


Db 5659 TTTGGAATGTTCTAAAAGGTGATATGAGCTGGTAGGAAACAGCTCCACCAACATTGGATG 5718
Qy 1153 AATTGGAATAATACCTCTGCTCAAAAGAGACCAATTGAGTTTTAAACCAGGATTACAG 1212
Db 5719 AGTACGAATCTTATACACCGGAAACAAAACGTCGCCCTCAGCTTTAAACCAGGATTACTG 5778
Qy 1213 GTCTCTGCGAGGTAGTGGTGGTAAATATATCAGAGACTTCGACGAGCTAGTTCGGTTGG 1272
Db 5779 GTCTTTGGCAAGTAAGCGGTGCAAGTGAATTAATCTGATTTTGATGAAGTTGTAAACATAG 5838
Qy 1273 ACTTAGCATACATTGATAAATTGGACTATCTGGTTCAGATATTAATAAATTTTAAAGACAG 1332
Db 5839 ACCTTGCTTATTTGGACGATGGACAATCTGGCGGATATCAAAATCTTATTTGAAACAA 5898
Qy 1333 TGAAGTTGTTATTTGAGAGAGGGAAGTAAGTAAAGTATATGAAAGTTTGGTTGGTCG 1392
Db 5899 TTAAGTAGTAGTAATGAAGGATGGAGCAAGTAGTGGCTTTCCACCATTTCTTTTAATGG 5958

RESULT 24
ABA01435
ID ABA01435 standard; DNA; 14142 BP.
AC ABA01435;
XX
XX 21-FEB-2002 (first entry)
XX
XX Streptococcus thermophilus eps5 operon #1.
XX
XX Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
XX ds.
XX
XX Streptococcus thermophilus.
XX
XX WO200179500-A2.
XX
XX 25-OCT-2001.
XX
XX 18-APR-2001; 2001WO-FR001199.
XX
XX 18-APR-2000; 2000FR-00004972.
XX
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX (DANO-) CIE DANONE SA GERVAIS.
XX (RHOD) RHODIA CHIM.
XX
XX Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX
XX WPI; 2002-017616/02.
XX
XX New nucleic acid fragments containing exopolysaccharide operon, useful
XX e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.
XX
XX Claim 11; Page 48-52; 144pp; French.
XX
XX The present sequence is an eps operon from Streptococcus thermophilus.
XX Proteins encoded by the eps operon function in exopolysaccharide (EPS)
XX synthesis. The operon is useful for producing chimeric eps operons, for
XX optimising production of EPS in lactic acid bacteria. EPS impart texture,
XX mouth feel and rheological properties to fermented food products (e.g.
XX yoghurt). They function as thickeners, to provide free-flowing and creamy
XX texture, and may also have biological activities beneficial to health
XX
XX Sequence 14142 BP; 4483 A; 2605 C; 2582 G; 4411 T; 0 U; 61 Other;
XX

Query Match 5.0%; Score 347.2; DB 6; Length 14142;
Best Local Similarity 55.5%; Pred. No. 7.2e-38;
Matches 766; Conservative 0; Mismatches 588; Indels 26; Gaps 4;
Qy 13 AATTGGCAATTTTGAATATGATAGAGTTGCAATTTCTGCAATCTTAAAGTCAATATAC 72
Db 4576 AAATGGTATATACAGTTGGTTGGTTGTTTTCGACGCCATGGTAGTAAATAAC 4635

Qy 73 CAAATGCTGATTAAATCGTTCTGGAATTTTTTATCATATGATGGTTTCATTTATTTTCAT 132
Db 4636 CTTATACAGAGATTACCAAGGAAGCAATTGTCCTTTTAGGTGTACATATGATGTCTTT 4695
Qy 133 TTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGGTAATCTGATAGAGTTGAAA 192
Db 4696 TCTATATCAGTAGTTTATATGAAAACTTAAAGTATAGAGGCTACTTGGATGAATCATTTG 4755
Qy 193 AAACATTTAACTATAGTATTAATTTTGGCAATTTTTCTTAGCGAGATATCAATTTTGTGG 252
Db 4756 CAACTGTCAAATATTTGTTTCATATTTGCTCTAAATGCAACATTTCTCGTTTTTTTCAG 4815
Qy 253 AGAATAATTTTCGACATTTCAAGACGTTGTCGCGTATTTTACATTAATAAATTCGTTT 312
Db 4816 ATGGAAAGTTTTTCAATCTCAGTCGCGGACTTCTTTACGTCACCTGATTTTCCAGTGTTC 4875
Qy 313 TGTATATACCTATTTAAACGTAATTTAAGCAGTTTTAAGGATAGCTTTCTATTTTTCGCAA 372
Db 4876 TCTTATACGTTACAAATACCTGTTCTTAAAGTATTTCCGTTTCATCTATTTATACAGCTCGTA 4935
Qy 373 TCTATCAAAAAAGAGCAGTTCTTAATTTACAAAGGCTGAACGATGGGAAAAATATGCAAGTTT 432
Db 4936 AAAGTAACACAGAAATATTTCTCTGATTTCTGATCAAGCACGCTCTAGAAAAT----- 4985
Qy 433 TATTTGAATCACATAAACAAATTTCAAAAAATCTTGTTCATTTGTCATTTGTTAGGTACAG 492
Db 4986 --GTTTTGCTCTGATGAAGACAATATGATGGTAGGATTACAGCAGTTTGTGCTTTGG 5043
Qy 493 AAATAGATAAAATTAATTTTATCAATTTACCGCTCTTATTAATTTCTTGGGAAGAGCTATAGAGT 552
Db 5044 ATAATCTTATTTACGGATCCATTTATCAAGAGTGTAAACCTGAAAAATTTGATTTGAAT 5103
Qy 553 TTTCAACAGGGAAGTGGTCGACAGCTCTTTATAATCTACCAAGTGAAGTTTGTAGAGC 612
Db 5104 ATGCGACACACTCAGTAGTAGACCAAGTTTGTATTAATCTGCCAAAGTGAAGTACAGA 5163
Qy 613 TAAAGCAATTCGTTTCAGATTTTGGAGTTGTTAGTATTTGATTAAGCGTTGATATTAAT 672
Db 5164 TTTGGGATTAATGCAATCACCTTTTGAAGTTAAGGATTCAGATATCCATTAATTTGAATG 5223
Qy 673 CATTCGGTTTTACTGCGTTTGAAAAAACAACAAATTTCAACTGCTAGGTGACCATAGCATG 732
Db 5224 CACTTGAATTTATAGATGAAGGTGAAAAGCGTATCCAAACAATTCGGTCTTTCAAAGTTG 5283
Qy 733 TAACTTTTCCACAAATTTTATAAGCTAGTATCATATCATATGATGAAGACGACTTTTGATA 792
Db 5284 TTACGTTTTTCAAGCGATTTTATAGCTATGAGATATCTTTGGCGAAACGTTTCTCTGATA 5343
Qy 793 TACTCGGAGCGGTAGTCGGGTAAATTTATTTGCTATAGTTTCTATTTTGTAGTTCCAA 852
Db 5344 TCTGTGAGCCCTAGTTTGGTCTCTGTGGATTTGGAATCTTCTTTATCCAC 5403
Qy 853 TTATTCGTAGAGATGGTGACCGGCTATTTTCTCAGAAACGAGTTTGGACAGAAATGGAC 912
Db 5404 TTATTCGTAGGATGGTGGCGCAGCCATTTTCTCAAGACCGTGTGGGAAAAATGGAC 5463
Qy 913 GCATATTTACATTTCAAGTTTCGATTCGATGATGTTGATGCTGAGGAGCGGAAAAAG 972
Db 5464 GTATCTTCAAGTTTATAAATTCGTTCTATGTGTGTTGATGGGAAAAATCAAGAAGA 5523
Qy 973 ACTTGCTCAGCCAAAACAGATGCAAGGTGGGTATGTTTAAATGCGGAAAAACGATCC 1032
Db 5524 ATTTGATGGCACAGAATCAAAATGCTGGTGGTATGTTT-----AAGATGGACAATGATCC 5578
Qy 1033 TAGAATTTACTTCCAAATTTGACATTTTATACGCAAAAAACAAGTTTAGACGAGTTACCAAGT 1092
Db 5579 ACGTATTTACCAAAATTTGGACATTTCAATTCG-THAAACGAGTCTTGATGAATTTCCACATT 5637
Qy 1093 TTTATTAATGTTTTAAATTTGGCGATATGAGTCTAGTTGGTACAGCTCCACCTTACAGTTGATG 1152
Db 5638 TTGA-----TGTTAGGTGATATGAGCTTTGGTAGGAACACGCTCCACCAACATTGGATG 5689
Qy 1153 AATTTGAAAAATATATCTCTCTGGTCAAAAGAGACGATTGAGTTTTTAAACCAGGGATTACAG 1212

Db 5690 AGTACGAATCTTATACACCGGAACAAAACGTCGCTCAGCTTTAAACCAAGTATTACTG 5749
Qy 1213 GTCTCTGCGAGGTAGTGGTCTAGTAATATACACAGACTTCGACGACGTAGTTCGGTTGG 1272
Db 5750 GTCTTTGCGAAGTAAGCGTCAAGTGAATTAAGTATTTGATGAAGTGTGAAGACTAG 5809
Qy 1273 ACTTAGCATATGATGAATTTGGAATCTCTGTCAGATATTAATAATTTTAAAGACAG 1332
Db 5810 ACCTTGCTTAATTTGACGAGTGAACATCTGCGCGATATTAATAATCTTATTGAACAACAA 5869
Qy 1333 TGAAGTTGATTTGAGAGGCGAAGTAAGTAAAGTATATGAAAGTTTGTGTTGTCG 1392
Db 5870 TAAAGTAGTAGTAATGAAGGATGGACCAAGTATGTCCTTCTCCATTTCTTTTAATGG 5929

RESULT 25
ABA01433
ID ABA01433 standard; DNA; 12786 BP.
XX AC ABA01433;
XX AC
XX 21-FEB-2002 (first entry)
XX DE Streptococcus thermophilus eps3 operon #1.
XX KW Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
XX KW ds.
XX OS Streptococcus thermophilus.
XX PN WO200179500-A2.
XX PD 25-OCT-2001.
XX PF 18-APR-2001; 2001WO-FR001199.
XX PR 18-APR-2000; 2000FR-00004972.
XX PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX PA (DANO-) CIE DANONE SA GERVAIS.
XX PA (RHOD-) RHODIA CHIM.
XX FI Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX WI MPI; 2002-017616/02.
XX PT New nucleic acid fragments containing exopolysaccharide operon, useful
XX e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.
XX PS Claim 11; Page 39-43; 144pp; French.
XX CC The present sequence is an eps operon from Streptococcus thermophilus.
XX CC Proteins encoded by the eps operon function in exopolysaccharide (EPS)
XX CC synthesis. The operon is useful for producing chimeric eps operons, for
XX CC optimising production of EPS in lactic acid bacteria. EPS impart texture,
XX CC mouth feel and rheological properties to fermented food products (e.g.
XX CC yoghurt). They function as thickeners, to provide free-flowing and creamy
XX CC texture, and may also have biological activities beneficial to health
XX SQ Sequence 12786 BP; 4161 A; 2228 C; 2047 G; 4055 T; 0 U; 295 Other;

Query Match 4.9%; Score 346; DB 6; Length 12786;
Best Local Similarity 57.2%; Pred. No. 1.1e-37;
Matches 727; Conservative 0; Mismatches 529; Indels 16; Gaps 5;

Qy 123 TATTTTGCATTTTATATCTCGTATGCCAGTTGAATTTGATGATAGAGGTAATCTGATA 182
Db 4649 TANGTGTCTTACTATATCAGTAGTTATTAATGAATCTTAAGTATAGAGGCTACTGGAT 4708
Qy 183 GAGTTTGAACAAACATTTAACTATAGTATAATATTGCAATTTTCTTACGCGAGTATCA 242
Db 4709 GAACTCATTTGCAACTGTCAATATTTGTTTCATATTTGCTCTTAATTGCAACATTTCTCTCG 4768

Qy 243 TTTTGTGAGAAATAATTTCCGCACTTTCAAGACGTGGTGGCTGTATTTTCCACATTAATA 302
Db 4769 TTTTTCGAGATGGAAGTTTTTCAATCTCACGTGCGGACTTCTTTAGTCCACCATGATT 4828
Qy 303 AACTTCGTTTGGTATACCTATTTAAAGTAATTTAAAGCAGTTTAAAGTAGCTTTCTA 362
Db 4829 TCAGGTGTTCTTTATACGTTTCAAAATACTGTTCTTAAGTATTTCCGCTCATTTAT 4888
Qy 363 TTTTCGCAATCTATCAAAAAAGACGATTTCTAATTAACAACGGCTGAACGATGGAAAT 422
Db 4889 ACAGCTGTAAGTAACAGAAATATTTCTCTTGATTTCTGATCAGGACGCTTGTGATAT 4948
Qy 423 ATGCAAGTTTATTTGAATCAATAAACAATTAACAAAAATCTTTGTCATTTGGTAGTT 482
Db 4949 GTTNTTATCTCGTATGAAGGACAATATGATGGTAGGA-----TTTCAGCAGTTTGT 5000
Qy 483 TTAGGTACAGAAATAGATAAAATTAATTTATCATTTACGCTCTATTTCTGTGGAAGAA 542
Db 5001 GTCTTGGATATCTCTTATTTTCACTGATCCATTTTATCAAGAGTGTAAACCTGAAATTTN 5060
Qy 543 GCTATAGAGTTTTCACAAGGGAAGTGTGCACACGCTCTTTTATAAATC-TACCAAGTGA 601
Db 5061 GATTGAATATGCCGACACACTCNAGTAGTACCAAGTTTGTGATTAATCNTGCCAAGTGA 5120
Qy 602 GTTTTATAGCTAAGCAATTCGTTTCAGATTTTTCAGTTTGTAGTATTTGATTAAGCGT 661
Db 5121 GCAGTACAAGATTTGGGATTTATGCTCACCATTTTGAACCTTATGGGATCCCAAGTATCAT 5180
Qy 662 TGATATTAATTCATTCGGTTTTTA-CTGGTTGAAAAACAACAAAAATCCAACTGCTAGTGT 720
Db 5181 TAATTTGAATGCCCTTGAATTTTATGAGTCAGGTCAGGTCGAAAAACATATCCAAATTTGGGTC 5240
Qy 721 ACCATAGCATTTGTAATCTTTTCCACAATTTTATAAGCCTAGTCATATCATGATGAAC 780
Db 5241 CTTTCAAGTTGTACGTTTCTAGGCAATTTTATAGCTATGGAGATATCTTGGCAATC 5300
Qy 781 GACTTTTGGATATATCTCGGAGCGGTAGTCGGGTTTAAATTTTGTGGTATAGTTCTATTT 840
Db 5301 GCTTCTCTCGATATCTGTGGAGCTCTAGTTGGTTGGTGTCTGTGGTATTTAGGAATCT 5360
Qy 841 TGTAGTTTCCAAATTTATTCGTAGAGATGGTGACCGGCTATTTTCTCAGAAAAAGTGTG 900
Db 5361 TCCTTTATCTCTTATTCGTAAGATGGGACACAGCTATTTTGTCTCAATACCGTGTGG 5420
Qy 901 GACAGATGGACGATATTTACATTTACAGTTTTCGATCCGATCTATGTCGATCTGAGG 960
Db 5421 GTGAAATGGTCGTATCTTTAAGTTTATAAATTCGTTCTATGCGTGTGATGCGGAAG 5480
Qy 961 AGCGCAAAAAAGACTTGTCTCAGCCAAAAACAGATGCAAGGTCGGGTATGTTTAAAAATGG 1020
Db 5481 AATCAAGNAGAAATTTGATGGCAGAGATCAATGTCTGGTGTATGTTT-----AAGAT 5535
Qy 1021 GAAAAACGATCTCTAGAAATTTACTCCAAATTTGGAATTTTATACGCAAAAAACAAGTTTACG 1080
Db 5536 GGCAATGATCCAGTATTTACAAAAATTTGACATTTTCAATCG-TAAAAACAAGTCTTGATG 5594
Qy 1081 AGTTACACAGTTTATAATGTTTAAATTTGCGGATATGAGTCTAGTTGGTACACGTCAC 1140
Db 5595 AACTTCCACAATTTTGAATTTTCTTAAAGGTGATATGAGCTTGGTATGGGACACGTCCTC 5654
Qy 1141 CTACAGTTGATGAATTTGAAAAATATATCTCTCGTCAAAAAGAGACGATTTGAGTTTAAAC 1200
Db 5655 CAACAGTTGATGATATGAAAAATATACCTGAAACAGAAACGTCGTTAAGTTTAAAC 5714
Qy 1201 CAGGATTTACAGGTCCTCTGGCAGGTTAGTGGTCGTAGTAATPATCACAGACTTCGACGACG 1260
Db 5715 CTGGTATCAACAGGCTTTTGGCAAGTAAAGTGGACGAAGTGAATCACTGACCTTTGATGAGG 5774
Qy 1261 TAGTTCGGTTGGACCTTAGCATAFACATTTGATTAATTTGCACTATCTGTCAGATATTTAAATTT 1320
Db 5775 TTGTTAAACTGGATGTTGCTTATATTTGATGGTGGCAAACTATGCGCGAGATATTCAAATCT 5834

| | | | |
|----|------|-------------------------------------------------------------|------|
| Qy | 1321 | TATTAAAGACAGTGAAAGTTGTATTGTTGAGAGAGGGAAGTAAGTAAAGTATATGAAAG | 1380 |
| Db | 5835 | TACTGAAACCGATTAAAGTAGTAGTAATGAAGGATGGAGCAAGTGATGGCTTTCACCAT | 5894 |
| Qy | 1381 | TTTGTTTGGTCG | 1392 |
| Db | 5895 | TTCTTTTAATGG | 5906 |

Search completed: April 29, 2005, 02:36:47
Job time : 3285 secs


```
;; SOFTWARE: <Unknown>
;; CURRENT APPLICATION DATA:
;;   APPLICATION NUMBER: US/09/107.433
;;   FILING DATE: 30-Jun-1998
;; PRIOR APPLICATION DATA:
;;   APPLICATION NUMBER: 60/ 085131
;;   FILING DATE: May 12, 1998
;;   APPLICATION NUMBER: 60/051553
;;   FILING DATE: July 2, 1997
;; ATTORNEY/AGENT INFORMATION:
;;   NAME: Ariniello, Pamela Deneke
;;   REGISTRATION NUMBER: 40,489
;;   REFERENCE/DOCKET NUMBER: GTC-011
;; TELECOMMUNICATION INFORMATION:
;;   TELEPHONE: (781)893-5007
;;   TELEFAX: (781)893-8277
;; INFORMATION FOR SEQ ID NO: 2303:
;;   SEQUENCE CHARACTERISTICS:
;;     LENGTH: 1443 base pairs
;;     TYPE: nucleic acid
;;     STRANDEDNESS: double
;;     TOPOLOGY: circular
;;     MOLECULE TYPE: DNA (genomic)
;;     HYPOTHETICAL: NO
;;     ANTI-SENSE: NO
;;     ORIGINAL SOURCE:
;;       ORGANISM: Streptococcus pneumoniae
;;       FEATURE:
;;         NAME/KEY: misc feature
;;         LOCATION: (B) LOCATION 1...1443
;;         SEQUENCE DESCRIPTION: SEQ ID NO: 2303:
US-09-107-433-2303

Query Match      5.9%; Score 415.8; DB 4; Length 1443;
Best Local Similarity 58.3%; Pred. No. 5.5e-66;
Matches 789; Conservative 0; Mismatches 552; Indels 12; Gaps 3;

QY 14 ATTGGCATATTATGATATGATAGCAGTTGCAATTTCTGCAATCTTAACAAGTCATATACC 73
DB 102 ATTGGCCATATGTCAGAGTTTCTTGTATTTATTTATGCTTATCTACTTAGCACTGTGAG 161

QY 74 AAATGCTGATTTAATCGTCTCGAATTTTATCATATGATGTTTCATTTATTTGCAAT 133
DB 162 AGAACAGAGATTGTTTCAACAACAGCTATTTGTAATTTATCTTCCACTATTTTGCCCT 221

QY 134 TTTTATATCTCGTATGCCAGTTGAATTTTGAGTATAGAGGTAATCTGATAGAGTTTGA 193
DB 222 TTATATCAGTATGATGACAGGATTCTTTAAAGGGGATATTTGATTGAATTTGTTCCA 281

QY 194 AACATTTAATATAGTATATATTTGCAATTTTCTTACGGCAGTATCATTTTGTGGA 253
DB 282 GACATTTGAATATATATCTTCTTTGGGCTAGCGATTAGTATTTCTAATTTTCTTTAGA 341

QY 254 GAATAATTTGCACTTTTCAAGACGTTGCGCTGATTTTACATTAATAAATTCGTTT 313
DB 342 AGATCGAATTTAGTATTTCCAGAGAGGCGATTTACTTCTCTCTATATATGTTCTCTT 401

QY 314 GGTATACCTATTTAACGTAATTTAAGCAGTTTAAAGGATAGCTTTCTATTTTCGCAAT 373
DB 402 AGTCTATGCTAAACCGATTATCAAGTGTATGGAAGGGGCTTATCCCACTTTAA 461

QY 374 CTATCAAAAAAGACGATTCTTAATTAACAAGGCTGAACGATGGGAAATATGCAAGTTT 433
DB 462 AGGAAGTAAGAAGATTCTCTACTTACAGCAACTTCTCGTGTGCAAAAGGATTGGATAG 521

QY 434 ATTTGAATCATATAACAATTTCAAAAATCTTTGTCATTGTTAGTTTGTAGTACAGA 493
DB 522 ACTAATAGATCAGATGATGTTGTTGGGAGTTGGTAGCCGTCAGTGTTTTAGATAAAC 581

QY 494 AATAGATAAAATTAATTTATCATTAACCGCTCTATTTATTTCTGTGGAAGAGCTATAGAGTT 553
DB 582 AGATTTTTCAGCATGATTTATTAAGGTTGTAG-----CAGAGGGGAGATCGTAACCTT 635
```

```
QY 554 TTCAACAAGGGAGTGGTCGACCAACGCTCTTTATAAATCTACCAAGTCAGTTTTTAGACGT 613
DB 636 TCGCACTCATGAGGTGGTCGATGAAGTCTTTATCAATCTTCCAAAGTGAAAAATACAAATAT 695

QY 614 AAAGCAATTCGTTTTCAGATTTTTCAGTGTGTAGTATTTAGTATGTAAGCCTTTCATATTAATTC 673
DB 696 TGGAGAGCTTGTCTCTCAGTTTGAACAGCATGGAGTGTGATTAACAGTCATCTAATATGC 755

QY 674 ATTCGGTTTTCAGTGGCTTGAATAAACAATAATCCAACTGCTAGGTGACCATAGCATTTGT 733
DB 756 TTTTCGATTTGATCTTGGCACATAAACAAGCAAAATTTGTGAGATGGCAGGACTAAACGTTGT 815

QY 734 ACTTTTTCACAAATTTTATAAGCCTAGTCATATCATCATGATGAACGACTTTTGGATAT 793
DB 816 GACTTTTTCACAAATTTTATAAGACTAGCCATGTGATTTGCTTAAGCGGGTTATTGATAT 875

QY 794 ACTCGAGCGGTAGTCGGGTTAATTTATTTGTGTATAGTTTCTATTTTGTAGTTCCAAT 853
DB 876 TATCGGTTCCCTGTTAGTTTGTATATGTTGTGTCTAGTCAGTATTTGTACTGGTTCCTTT 935

QY 854 TATTCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAACGAGTTTGGACAGATGGAACG 913
DB 936 GATTCGAAAGGATGGGGCTCTGCTATTTTGTCTCAGACGCTATTTGGGAAAAATGTCG 995

QY 914 CATATTTACATTTCTACAAGTTTCGATCGATGATGTTGATGCTGAGAGCGCAAAAAAGA 973
DB 996 CCAATTTCACTTTTACAAGTTTCGCTCTATGTTGTGTGATGCTGAGGAGAAAAAGAGA 1055

QY 974 CTTGCTCAGCAAAACACAGATCAAGGGTGGGTATGTTTAAATGGAAAAACGATCCT 1033
DB 1056 ACTCATGGAACAAANATACCATGCA-----GGGTGGAATGTTTAAGGTGGATGAGATCCA 1110

QY 1034 AGAATTAATCTCAATTCGACATTTTATACGCAAAACAAGTTTATAGACGAGTTTACCACAGTT 1093
DB 1111 CGTATCACGAAATTTGCTCATTTTATACG-GAAGACGAGCTTGGACGAGCTTACCACAGTT 1169

QY 1094 TTATATGTTTAAATTTGGCGATATGATCTAGTTTGTGTACACGCTCCACCTACAGTTGATGA 1153
DB 1170 TTACAATGTTTAAAGGGAGATATGATTTGGTAGGACACGACCAACAGCTGGACGA 1229

QY 1154 ATTTGAAAAATATACCTCTCTCAAAAGAGACCATTTAGTTTTTAAACACAGGATTTACAGG 1213
DB 1230 GTATGAGCACTATACCCGACAAACAAACGTCGCTTAAGTTTAAACCTGGCATACAGG 1289

QY 1214 TCTCTGGCAGGTTAGTGGTCTGATGTAATATACAGACTTCGACGACGTCAGTTTCGTTTGA 1273
DB 1290 TCTATGGCAGGTCAGCGGACGAAGCGAGATCAAGAATTTTCGATGAGGTTGTCAAATTTAGA 1349

QY 1274 CTTAGCATATGATGATTAATTTGGACTATCTGCTCAGATATTAATAATTTTATTAAGACAGT 1333
DB 1350 TGTGGCTATATAGACGGTTGGCAATTTGGAAAGATATTTGAAATTTTATTGAAGACAGT 1409

QY 1334 GAAAGTTGTATTCTTCAGAGAGGGAAGTAAGTA 1366
DB 1410 TAAAGTTGTATTGATGAGGATGGAGCGAAGTA 1442
```

RESULT 4

```
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
```


ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 2.3%; Score 158.4; DB 4; Length 1830121;
Best Local Similarity 56.4%; Pred. No. 6.7e-19;
Matches 359; Conservative 0; Mismatches 266; Indels 11; Gaps 3;

Qy 747 AATTTTATAGCCCTAGTCATATCATGATGAACGACCTTTGGATATATCTCGGAGCGGTA 806
Db 924401 AATCTGCTAAACTATCTCTCGTATTTTGAACGACCATGATATTTGTTGGCTCC 924460

Qy 807 GTGCGGGTAAATTTATTTGGTATAGTTCTATTTTGTAGTTCCAAATTTATTCGTAGAGAT 866
Db 924461 TTAGCTATTTATTTATTTTCCCGAGTGTCTTATCTGTATTTTGGCAGTAAAAAAGAT 924520

Qy 867 GTGCGCGGCTATTTTGTCTCAGAAAGAGTGTGGACGAATGACGACATTTTACATTC 926
Db 924521 GGTGTAATGCTATTTATTTATGGGCATCTAGAAATGCTCGTAAATGAAAAAACAATTTAACTGT 924580

Qy 927 TACAAGTTTCGATGATGTATTTGATGCTCAGAGCGCCAAAGAGCTTCTCTCAG---- 982
Db 924581 TTGAATTTAGAACAAATGGCTGTGAATTTCTAAAGAGTGTAGATGAATTCCTAAGAAC 924640

Qy 983 ---CCAAACACGATGCAAGGGTGGGTATGTTTAAATGGAAAAACGATCCTAGAAAT 1039
Db 924641 GATCCTGAGCAAGAGCTGAATGGAAAAAGATTTTAAATTAAGATGATCCTAGAAAT 924700

Qy 1040 ACTCAAATTGACATTTTCATACGAAACAAAGTTTACGAGTTTACACAGTTTATATAA 1099
Db 924701 ACTAAATCGGTGCTTTTATTCGC-AAAAAAGTTTATAGATGAATTTACCTCAATTTTAA 924759

Qy 1100 TGTTTTAAATGGCATGATGCTAGTGTGTTGATGCTCAGCTCCACCTAGATTTGA 1159
Db 924760 TGTTTAAAGGTGAATGATTTAGTTGGCCCTCGCCCAATTTGAATTTGATGAGCTAGA 924819

Qy 1160 AAAATATCTCTCTGTCAAAAAGACGATTCAGTTTAAACGAGGATTAACAGTCTCTG 1219
Db 924820 GCGTTATGAAGAAACGATAGATTATTAATTCGACAGACCCGATGACAGGGCTTTG 924879

Qy 1220 GCAGGTAGTGGTTCGTAGTAAATATACACAGACTTCGACGACGTAGTTCGGTTGGACTTAGC 1279
Db 924880 GCAAGTCAGCGGACGTAATAATATC---GATTATAATACTCGCGTTTATTTGATTTCTTG 924936

Qy 1280 ATACATTGATTAATTTGGACTATCTGGTCAGATTTTAAATTTTATTAAGACAGTGAAGT 1339
Db 924937 GTATGTAATAAATTTGGTCACTTTGGAAATGATTAATTCGATTTTATTTAAACAATGAATGT 924996

Qy 1340 TGTATTGTTGAGAGGAGGAAGTAAGTAAAGTATATAT 1375
Db 924997 GGTTTAAATCGTATGAGCTTATTAACGTTATTT 925032

RESULT 5
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 2.3%; Score 158.4; DB 4; Length 1830121;
Best Local Similarity 56.4%; Pred. No. 6.7e-19;
Matches 359; Conservative 0; Mismatches 266; Indels 11; Gaps 3;

Qy 747 AATTTTATAGCCCTAGTCATATCATGATGAACGACCTTTGGATATATCTCGGAGCGGTA 806
Db 924401 AATCTGCTAAACTATCTCTCGTATTTTGAACGACCATGATATTTGTTGGCTCC 924460

Qy 807 GTGCGGGTAAATTTATTTGGTATAGTTCTATTTTGTAGTTCCAAATTTATTCGTAGAGAT 866
Db 924461 TTAGCTATTTATTTATTTTCCCGAGTGTCTTATCTGTATTTTGGCAGTAAAAAAGAT 924520

Qy 867 GTGCGCGGCTATTTTGTCTCAGAAACGAGTGTGGACGAATGACGACATTTTACATTC 926
Db 924521 GGTGTAATGCTATTTATTTATGGGCATCTAGAAATGCTCGTAAATGAAAAAACAATTTAACTGT 924580

Qy 927 TACAAGTTTCGATGATGTATTTGATGCTCAGAGCGCCAAAGAGCTTCTCTCAG---- 982
Db 924581 TTGAATTTAGAACAAATGGCTGTGAATTTCTAAAGAGTGTAGATGAATTCCTAAGAAC 924640

Qy 983 ---CCAAACACGATGCAAGGGTGGGTATGTTTAAATGGAAAAACGATCCTAGAAAT 1039
Db 924641 GATCCTGAGCAAGAGCTGAATGGAAAAAGATTTTAAATTAAGATGATCCTAGAAAT 924700

Qy 1040 ACTCAAATTGACATTTTCATACGAAACAAAGTTTACGAGTTTACACAGTTTATATAA 1099
Db 924701 ACTAAATCGGTGCTTTTATTCGC-AAAAAAGTTTATAGATGAATTTACCTCAATTTTAA 924759

Qy 1100 TGTTTTAAATGGCATGATGCTAGTGTGTTGATGCTCAGCTCCACCTAGATTTGA 1159
Db 924760 TGTTTAAAGGTGAATGATTTAGTTGGCCCTCGCCCAATTTGAATTTGATGAGCTAGA 924819

Qy 1160 AAAATATCTCTCTGTCAAAAAGACGATTCAGTTTAAACGAGGATTAACAGTCTCTG 1219
Db 924820 GCGTTATGAAGAAACGATAGATTATTAATTCGACAGACCCGATGACAGGGCTTTG 924879

Qy 1220 GCAGGTAGTGGTTCGTAGTAAATATACACAGACTTCGACGACGTAGTTCGGTTGGACTTAGC 1279
Db 924880 GCAAGTCAGCGGACGTAATAATATC---GATTATAATACTCGCGTTTATTTGATTTCTTG 924936

Qy 1280 ATACATTGATTAATTTGGACTATCTGGTCAGATTTTAAATTTTATTAAGACAGTGAAGT 1339
Db 924937 GTATGTAATAAATTTGGTCACTTTGGAAATGATTAATTCGATTTTATTTAAACAATGAATGT 924996

Db 924581 TTGAATTTAGCAATGGCTGTAATTTCTAAAGAGTGTAGTGAATTCCTAAGACA 924640
Qy 983 ---CCAAACACAGATCAAGGGTGGGTATGTTTAAATGGGAAACGATCCTAGAAAT 1039
Db 924641 GATCCTGAAGCAAGAGCTGAATGGGAAAGATTTTAAATTTAAAGATGATCCTAGAAAT 924700
Qy 1040 ACTCAATTGGACATTTTCATACGCAAAACAGTTTATAGACAGGTACACAGTTTATATA 1099
Db 924701 ACTAAATTCGGTCTTTTATTCGC-AAACAAAGTTTAGATGAATTTACCTCAATTTATTTAA 924759
Qy 1100 TGTTTAAATGGCGATATGATCTAGTTGGTACAGTCCACCTACAGTTGATGAATTTGA 1159
Db 924760 TGTTTTAAAGGTGAATGAGTTTATGTTGGCCCTCGCCCAATGTTAATGATGAGCTAGA 924819
Qy 1160 AAAATATACCTCTGGTCAAAAGAGACGATTTAGTTTAAACACAGGGAATTACAGTCTCTG 1219
Db 924820 GCGTTATGAAGAAACGTAGATTATTACTTAATGGCAAGACCCGGTATGACAGGCTTTG 924879
Qy 1220 GCAGTTAGTGGTCTGATGATATATACAGACTTCGACGAGTAGTTCGGTTGACCTTAGC 1279
Db 924880 GCAAGTCAGCGGAGCTAATAATATC---GATTATAATACTCGCGTTTATTTTGTGATCTTG 924936
Qy 1280 ATACATTGATTAATGGGACTATCTGTCAGATATTAAATTTTATTAAGACAGTGAAAGT 1339
Db 924937 GTATGTAATAAATTTGGTCACTTTGGAATGATTTGCGATTTTATTTAAACAATGAATGT 924996
Qy 1340 TGTATTGTTGAGAGAGGGAAGTAAGTAAAGTATAT 1375
Db 924997 GGTTTTAAATCGTATGAGCTTTAATACGTTAATTT 925032

RESULT 6

US-09-107-532A-2527
; Sequence 2527, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2527:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...984
; SEQUENCE DESCRIPTION: SEQ ID NO: 2527:
US-09-107-532A-2527

Query Match 2.2%; Score 152.6; DB 4; Length 984;
Best Local Similarity 51.4%; Pred. No. 1.6e-18;
Matches 384; Conservative 0; Mismatches 354; Indels 9; Gaps 1;
Qy 4191 AAAATGAATGATTTAAATTCAGTTATTGTACCAATTTTAAATGTCCAAGATTATCTTGA 4250
Db 12 AAATAATAATGTGTGAGATTAGTATTATTTCTGTGTATAAAGTTCAAAAATATTTAAG 71
Qy 4251 TAAATGTATTAAACAGTATTATTAAACCAACATATATAATTTAGAGTTATCTTCGTA 4310
Db 72 AAAATGTAGATTTCGATTTTAGCCCAACATTTACTGATTTTGAAGTTATATATTAGTAG 131
Qy 4311 TGATGGAAGTACTGATGATTTCTGAGAAAATTTGCTTAAACTATATGAAGACGATGGAAG 4370
Db 132 TGATGATCTCTCGACAAATAGCGGGAAAATCTGTGATGATGATGCTGAAAAGATATATCG 191
Qy 4371 AATTAAATATTACAAGAAAATTAATCGCGTCTPAGCAGATGCTCGAAAATTTCCGACTAGA 4430
Db 192 AGTACGTGTGATCCATTAAGAAAACCGCGGATTTAGCAGTGTCTAGAAAATCGCGGGATTGA 251
Qy 4431 ACATGCAACAGGTAAATATATTGCTTTTGTTCGATTTCTGATGACTATATAGAAGTTGCAAT 4490
Db 252 TGTTCGAGAGGCAATATCTTAGGCTTTTGTGACAGTGTATATATTGATGAAGATAT 311
Qy 4491 GTTCGAGAGATGCATGATAATAATACTGATGATTAATGCGGATATAGCAGATAGATTTT 4550
Db 312 GTATGAATTTTGTATGAGAAATTTGAAAATTTCAATGATGCCGCAATTAGCTCTGTGAAAT 371
Qy 4551 TTCTTTAGTAGAGCAAAACGGGTATACAAAGAAAAGAAAATAGTAATTTTTCATGCTTT 4610
Db 372 GATTCCTTTTATGTTGATAGATATTAAGAAAGCTATTAAGAAAAGAAAGTAAATATT 431
Qy 4611 AACGAGAGAGAGACTGTAAAGAAATTTTGTCCAGATCTTAATATAGAAAATTAATGTTG 4670
Db 432 AAATAAAAAGGAAGCAATAAAATCTGTTTAGAAGGAACGCAATTTTATGCATACGCATG 491
Qy 4671 GTGCAAGCTTTTATTCAGGAGATATTATAAGAGATATAAATTTCCAAATTAATATAGAAG 4730
Db 492 GAATAACTTTATCGTAAGAACTTTTAAAGATAATCGTTACTTGGATGGAATAAATTT 551
Qy 4731 TATTGTTGAGGATTTGCTTTTAAATTTGGAGGCTTTTGAACAATGTAAACAGCTGTAGTAGT 4790
Db 552 TGAGGATGATATATAATATAGATCTATTTATTT-----TCAACACAGAAAATTTGT 602
Qy 4791 TGATACCTAGAGAAATATTATTATAATTTATGTCATTTCGTAACAGTTTCGCTTATTAATCAGAA 4850
Db 603 TGTTTCCAATGAAGAGAAATATTCTTCTTCAAGAAATGATAGTATTATTTGGAAAAG 662
Qy 4851 ATTCTCTATAAATAATATTGATTTAGTCACAGATTGGAGAAATACCCCTTTAAGTTAAA 4910
Db 663 TTTTCTATGAATAATTTTCGATGTTATGAAGCATGCGAATATAACAAGAGAAGATTTT 722
Qy 4911 AAGAGAGTTTAGTCATTATTTTTCATGTC 4937
Db 723 AGATGCTTTTCCAGATTTCATGATTC 749

RESULT 7
US-09-134-000C-2987
; Sequence 2987, Application US/09134000C
; Patent No. 6617156

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2987
LENGTH: 993
TYPE: DNA
ORGANISM: Enterococcus faecalis
US-09-134-000C-2987

Query Match 2.1%; Score 145.4; DB 4; Length 993;
Best Local Similarity 52.0%; Pred. No. 3.2e-17;
Matches 382; Conservative 0; Mismatches 341; Indels 12; Gaps 2;
Qy 6164 TATTAGTAAATTTCTATAATTTGACCTATATATATAGTAAATATTTATCTAAATG 6223
Db 18 TATGCCAAATTTAGTATTTTCTCTATACATGTAGAAATATTTAGAAATG 77
Qy 6224 TATAGATAGCATTTGTAATTCAGACCTACAAACATATAGAGATTTCTTCTGGTGAATGACGG 6283
Db 78 TGTACGCTCTATTTAGTCTCAACGTTTACTGACTTTTGAATTAATTTCTGGTGGACGATGG 137
Qy 6284 TAGTACGATTAATTCGGAAGAAATTTGTTTGTAGCATATCGGAAGAAAGATAGTCGATTCG 6343
Db 138 CTCTCCAGACAGTTCTCGGAGCAATGTGTGATCAGTTCGTCTGAACAGATCAACGGGTAA 197
Qy 6344 TTATTTTAAAGAGAACGGCGCTATCAGATCCGCTAATATGATGATGATGATGATGATGATG 6403
Db 198 AGTATCATTAAGAAATGTGGCTAAGCATGCTGTAATGCTGGAATGAAATAGC 257
Qy 6404 CAAGGGTACTTCTAGCTTTTATAGACTCAGATGATTTTATTCATTCGGAGTTCATCCA 6463
Db 258 AACAGGTGAGTATTAGTTCGTTAGATGATGATGATGATGATGATGATGATGATGATGATG 317
Qy 6464 ACGTTTACAGCAATTTAGAGAGAAATGCGCTTGTGGCAGTTGCTGTTGATGATGATG 6523
Db 318 ACTATTATATACAAATATCGTAAAGAGATGCGGATTTATCAATCTGTGTTTATGTA 377
Qy 6524 GSTAGATCTTTCGGGCAATTTCTTAACAGCAGCGCTTCTCAAAATCAGGCTGTCT 6583
Db 378 TGTCTATG----AGGGGAAAGAGCCCAATTTGTAAGATTTTAATACAGGGAACCTTTTCCA 433
Qy 6584 GAGCGGCAAGGAATGTTTCTAAAGAGCTGCTAGAGCGCGATGCTCATCGCTTTGTGGTGGC 6643
Db 434 GAGAAGAGCAATATTGTTAA-----TTTTACAGGAATATTATCTCGTACATGC 485
Qy 6644 CTGTAATAAATCTATATAAAGAACTATTGTAAGATTTTCGATTTGAAAGGGTAAGAT 6703
Db 486 TGTAAATAGTTATATAAAGAAATCTTTTGGCGATTTAAGATATCTTAAGGAAAGTA 545
Qy 6704 TCATGAGATGAATACCTTCACTTATCGTCTCTATGAGTTAGAAAGTTGCAATAGT 6763
Db 546 TCATGAGATTCATTTATATAGTATGATTTGTTGCTGGAATGTCAGAAAGTCTCTATTGA 605
Qy 6764 TAAGGAGTGTGTTGACTATTATGTTGACGAGAAATATAGTATCAAACTTCTAGCATGAC 6823
Db 606 TTAACACAAAATATTACTATTATCATCGAATGGGCGAGCAATTAATACAGAAACATTTTC 665
Qy 6824 TGACCATCGCTTCCATTCGCTACTGGAATTTCAAAATGAAGCAATGGACTTCTATGAAG 6883
Db 666 AGATAAACAAATTCAGTTTATAGAGCTTGGGAGAGAAATGAACTTAATTTGAAGGTAA 725
Qy 6884 TAGAGAGATTAAGA 6898
Db 726 GGGAGCAGTTATAGA 740

RESULT 8

US-08-961-527-71
Sequence 71, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 32768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-71

Query Match 2.1%; Score 144.6; DB 3; Length 32768;
Best Local Similarity 52.3%; Pred. No. 9.1e-17;
Matches 368; Conservative 0; Mismatches 329; Indels 6; Gaps 2;
Qy 5144 AGAAATTTCAAAACGACGTAGAGGTAAATAATGATAAAATTAAGTGTATTGTTCCAGTTTA 5203
Db 6153 ACATAAATCAAGGAGAAAATTCCTAGTGGATGATAAAATAACAGTCATTGTACCAGTATA 6212
Qy 5204 TAATGTAGATAAATTTTAAGTAGTTTGTATAGAAGCAATTAATTAATCAAAATTAATAAA 5263
Db 6213 CAATGTGGAACACTATCTGAGGAAGTGCCTAGATAGTATTATTACTCAACATATAAAAA 6272
Qy 5264 TATAGAAATATTAATGTATAGATGATGCTCTGTAGATGATTCTGCTAAAAATATGCAAGCA 5323
Db 6273 TATTGAGATTTGTCGTGTTAATGATGTTCTACCGATGCTTCAGGTGAAATTTGTAAGA 6332
Qy 5324 ATATCGAGAAAAGATAAAAGAGTAAAAATTTTTTTCATCAATCATAGTGAGGTATCAAA 5383
Db 6333 ATTTTCAGAAATGATCACCAGAAATTCCTATATAGAAACAGAAAATGCTGCTTTCTGTC 6392
Qy 5384 TGCTAGAAATCATGGAATTAAGCGGAGTACAGCTGAATATATATTATTTGTTGTTGACTCTGA 5443
Db 6393 CGCAGAAACACCGGCTCTGAATATATATGTCGGAAATTTATGTGACCTTTGTGGACTCGGA 6452
Qy 5444 TGATGTTGTTGATAGTAGATTAGTAAAGAAATTAATTTTAAATATTAATAAAGTAGAAG 5503
Db 6453 TGATTGGATTGAGCAAGATTATGTAGAACTCTATATAAATAAATAGTAGAGTATCAGGC 6512
Qy 5504 TGATTATCTGTTGTTGTTGACGCTACTTTTTCAGAAAAATATAAATAATTTTGAAGTCAA 5563

Db 6513 TGATATTGTCAGTGGTAAATTAATTTCTTTCAACGAAAGTGAAGAAATGTTCTACTTCA 6572
 Qy 5564 TA-ATCCAAATATGATTTTGAACCAATTAATACCGTGCAGGACATGGGAGAAAATTT 5622
 Db 6573 TATATTGGGAGACTCCCTATTATGAGAAAGTATATGATATGTTCTTCTATCTTTGAGAACTT 6632
 Qy 5623 TTATGAATTTGTATATATAAATAATTTTCTA-----CTCCGTTTGTAAACTATATAA 5677
 Db 6633 GTATGAACCTCAAGAAATGAAGAGTTTGTCTTGATATCTGCTGGGGTAACTCTATAA 6692
 Qy 5678 GAAAAGATACATAACAGACTCTTTTCAAGAGAAATCAATGGTTAGGAGAGATTTACTTTT 5737
 Db 6693 GGCAGAGATTGTTGAGCAGTTGCGCTTTGACATAGGTAATTTAGGAGAGATGTTACCT 6752
 Qy 5738 TAATCTGCATTTTAAAGAAATATAGATAGAGTTAGTTTATTTGACTGAACTTTTATTT 5797
 Db 6753 CAATCAAAAGGTATATTTATTTATCAGAAAAGGTAAATTTATTTAAATAAAGTCTTTATGC 6812
 Qy 5798 TTATAGGAGAGGTATATCTTAAGTACAGTAAATTTCTTTTAAAGAA 5840
 Db 6813 TTATCGGATTAGAAAAGGTAGTTTATCAAGAGTTTGACAGAA 6855

RESULT 9

US-09-134-000C-2344
 ; Sequence 2344, Application US/09134000C
 ; Patent No. 6617156
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 032796-032
 ; CURRENT APPLICATION NUMBER: US/09/134,000C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/055,778
 ; PRIOR FILING DATE: 1997-08-15
 ; NUMBER OF SEQ ID NOS: 6812
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2344
 ; LENGTH: 2529
 ; TYPE: DNA
 ; ORGANISM: Enterococcus faecalis
 US-09-134-000C-2344

Query Match 2.0%; Score 139.2; DB 4; Length 2529;
 Best Local Similarity 44.7%; Pred. No. 5.1e-16;
 Matches 677; Conservative 0; Mismatches 823; Indels 14; Gaps 3;
 Qy 4198 AATGATTTAATTTTCAGTTTATGTACCAATTTATTAATGTCGAAGATTATCTTGATAATGT 4257
 Db 25 AATGTGAAGTGTGAGTCAATGTACCGTCTATTAACGTAGAAACCTATTTAGAAGGCA 84
 Qy 4258 ATTAAACAGTATTAATTAACCAACATATACTAATTTAGAGGTTATTTCTCGTAAATGATGGA 4317
 Db 85 TTAATGATTTGAAAAACCAACATTAAGAGAAATTTGAATTTCTTAATTAATGATGCG 144
 Qy 4318 AGTACTGATGATTTCTGAGAAAATTTGCTTTAACTATATGAGAACGATGGAAGAAATTA 4377
 Db 145 TCTACTGATATAGTCAAAAATTTATTCAAGAAATTTGCAAAAGATGATCCACGTTTTCGC 204
 Qy 4378 TATTACAGAAAATTAATGCGGTCTAGCAGTCTCGAAATTTTCGGACTAGACATGCA 4437
 Db 205 GTTTTCAATGTTAAATGGGGAAATTTGGAAGAACCTTTTAACTTGGGTGTTTTCAGAAGCA 264
 Qy 4438 ACAGTAAATATATGCTTTTGTGCTGATTTCTGATGACTATATAGAAATTTGCAATGTTTCGAG 4497
 Db 265 AAAGTGAATATATGCTGAAATTTGAAAGTGAACGATTTATGATGATGCTGCTATGAG 324
 Qy 4498 AGAATGATGATATATACTGAGTATTAATGCCGATATAGCAGAGATAGATTTTGTGTTA 4557
 Db 325 CGTTTATACATACGGCTTAAAGTCATCATGACAGATGTTGTCGCTTGTGTAATTTGGGTTGAA 384
 Qy 4558 GTAGACGAAACGGGTATACAAAGAAAAAAGAAATAGTAATTTTCTATGCTTTAAACGAGA 4617

Db 385 TTTTCTTCTGAAGAAGAGGTTGAGCCGATATATTTATGGCAATATCCCGATAATAACAAT 444
 Qy 4618 GAAGAGACTGTAAAGAAATTTTGTGAGGATCTAAATATAGAAATTAATGTTGGTGCAG 4677
 Db 445 CAAATGATGACTTAAAA--ACGCGGATTTAAATTTGACAGGTTTATCCATGGAATGCC 501
 Qy 4678 CTTTATTTCCAGAGATATTAAGAATATAAATTTCCAAATTTCAATTAATATAGAGTATTG 4737
 Db 502 ATTATAAAAAAGATGATGTTGAGAAAGAAACGTAACTTGGGATGAGGAAATTTAAAGT 561
 Qy 4738 GAGGATTTGCTTTTAAATTTGAGGTCCTTGAACAATGTAACAGTGTAGTGTGATCT 4797
 Db 562 TATGGGACACTGGTCTATTTTGGAAAAATCAATTTAGCATCTCAGAAATGTTATTTTAT 621
 Qy 4798 AGAATAATATTAATTAATTAATGATCTTCGTAACAGTTCGCTTATTAATCAAGAAATCTCT 4857
 Db 622 AAAGATTGTTGTTATTTTATAGACAAGAACCCAAATTCGACAGTAATAATATGTCGA 681
 Qy 4858 ATAAATAATATTAATTTAGTCAACAGATTGAGAAATTTACCCCTTTAAAGTTAAAAAGAG 4917
 Db 682 ACAAGGTTCCCTTTCTT-GTTTCAACAATTTAAACTAAATTCGTTCAAAATTTAATGGA-- 737
 Qy 4918 TTTAGTCAATTAATTTGATGCAAAAGTTATTAAAGAGAAGGTTAAATGTTTAAACAAAATG 4977
 Db 738 -----ACAAATAAATTTGAACGCTCAAGGTTATTTTATAACAATAATGTTCCAAA 790
 Qy 4978 TATTCAACAGATTTGTTGGATTAATGATGTTCTTCCCAATATTAAGATCTTTATCGAAAAGAA 5037
 Db 791 AATACTTTTGGGCGATTGAAAAATTAACCCATTTACGTGATGATGTCGTTTATGAAGTGA 850
 Qy 5038 ATACGTAGATATCCATTTTAAAGCGAAAGATATTTATCAAGAAAGCATTTAGTTACG 5097
 Db 851 TTCAAAAGGTGGCAGTAGATTTTAGACAAGCTTTTAGAAAACAGACCAGCTGAATGATATCG 910
 Qy 5098 TTGTAATTTGATGAAATTTTCGCTAAACTATATGTAATGTTATATAGAAATTTTCAAAAG 5157
 Db 911 ATTTTGAATACATCAACAATTTTATCAAAATGCAATGATCCAGAGCTTATATGAGG 970
 Qy 5158 CAGTAGAGGTAAATAATGGAATAATTTAGTGTATTTGTTCCAGTTTATAATATAGATAAAT 5217
 Db 971 ACTACTTGAATAATCTGTATAAAGTTAGTGTGTCATGCGGATTCACAAATGCTTCAAAAT 1030
 Qy 5218 ATTTAGTAGTTGTATAGAAAGCATTTTAATCAAAATTTATAAAATATAGAAATATATAT 5277
 Db 1031 ATTTAAGACAAACATTTAGAAACCTGTTGTGAGCAGCTTTTACGTGAGATTTGAAATCATTT 1090
 Qy 5278 TGATAGATGATGCTCTGTAGATGATTTCTGCTAAATATGCAAGGAATATGCAAGAAAG 5337
 Db 1091 TAGTAGAGATGCTCTACAGATAATACCATGATATTTAAGCAATTTGCAAGTAAAG 1150
 Qy 5338 ATAAAGAGTAAATAATTTTTTTCATTAATCATAGTGGAGTATCAAAATGCTAGAAATCATG 5397
 Db 1151 ATCCGAGAAATACGGGATTTCTTATTTGGCAAAAGTAATCTCGACACGCAAGAAATGTGG 1210
 Qy 5398 GATAAAGCGGAGTACAGCTGAATATATATGTTTGTGACTCTGATGATGTTGTTGATA 5457
 Db 1211 GCATTTGATGCAACGAGGACGGTATTTGCAATTTCTGATGAGATGATCATTTTGAAG 1270
 Qy 5458 GTAGATTAGTAGAAAAATTTATTTTAAATATTAATAAAGTAGAAGTGAATTTATCTGGTT 5517
 Db 1271 CGAATTTTATCAAGACGCTTATTTATCGTGCAATATGATAGTGCAACNGATATTTCTGTAT 1330
 Qy 5518 GTTTGTCGCTACTTTTTTTCAGAAAAATATAAATAATTTTGAAGTGAATAATTTTCAAAATTTG 5577
 Db 1331 TTGGCATGAAAGAAAAGTTTACCATAATGGAGAGTGCACGTTGTTTCAAAATCCATTAATA 1390
 Qy 5578 ATTTTGAAGCAATTAATACCGTGCAGGACATGCGGAGAAAAAATTTTATGAAATTTGTATA 5637
 Db 1391 CTAATGCGGACGATATGCTGTTGGAAGAAATTTTCAATTTGATGAAGTAACCCCTTATTTAT 1450
 Qy 5638 TAAATAATATTTTTTCTACTCTCTCTGTTGTAACATATATATAAGAAAAAGATACATAACAGATC 5697


```

; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14164
; LENGTH: 18773
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14164

```

| Query Match | 1.9% | Score 132.6 | DB 4 | Length 18773 |
|-----------------------|----------------|------------------------------------------------------------------|-----------|--------------|
| Best Local Similarity | 47.2% | Pred. No. 1.2e-14 | | |
| Matches 604 | Conservative 0 | Mismatches 659 | Indels 16 | Gaps 6 |
| Qy | 2850 | AAGTATATAAAGTGGAAABAAATCAACATCTTCTCTATTGTCCTTACAGATGATTCAT | 2909 | |
| Db | 18379 | AATACATATATAAATATATATATAATAAAATTCCTATATAAAATATATATATATAT | 18320 | |
| Qy | 2910 | TGATGAATTCCTAAATGCAAGAAATTTAGGTTTTTTTATTTGCTAGAAAGCTAAAAATAGA | 2969 | |
| Db | 18319 | ATATAATATCTATATAAATA-----TATATGATATTATATATAATATCTATATA | 18268 | |
| Qy | 2970 | AAATAAATCTAAATTTAAAGAAATTAATCTAAAAATAAAATAGTTGATTTGTGAGAG | 3029 | |
| Db | 18267 | AAATATATGTATATTTATATAGAACTATATATATATAAATATATATATATAT | 18208 | |
| Qy | 3030 | TAATGATGTTTAAATTTAAATATGATGACCCGGATATTTTATTTTAAAGTACTCTGG | 3089 | |
| Db | 18207 | TATATATTAATTAATAATATATAATGATTTATATAAATAAATATATATAAATAT | 18149 | |
| Qy | 3090 | TTGATTTATTTTATTCAGAGCAAAAGTAGTAGTTTATTAATTTTATTTTATGATTTAAAT | 3149 | |
| Db | 18148 | TAATTTTATATTAATAATTTATATAAATTAATATAATAAATATATAATATATAAAT | 18089 | |
| Qy | 3150 | TTATTTTCATATAAAATTTTGAACCTAAGCTAAATATTAATAAAATGAAATTTTATGTTT | 3209 | |
| Db | 18088 | ATATATATATCATATGATATAAAATATAAATATATATAATATAATATATGTAATGT | 18029 | |
| Qy | 3210 | TTATATATGGTCTATATATGTTTTGTTTCAGTAGTCACAAGTAGTGTGTTGGAATAAT | 3269 | |
| Db | 18028 | ATAATATATATAATATATATATATATATATATAAATAATATAAATATATATATAA | 17969 | |
| Qy | 3270 | TTTGAAGAAGTATTTGCGAGAT-TTTACTGCTCCCATAAATTTGCGATTTTGCAATATGTA | 3328 | |
| Db | 17968 | TAGTATATATAAATTTCAATATATATAAATAATATCATATATAATATAATATAAATA | 17909 | |
| Qy | 3329 | TTATAATTTGTATTTCAATTTATAAATATGATTTATAAAAAATTTAAAAATAGTATCTTTT | 3388 | |
| Db | 17908 | ATATATATCATATAATATAATATATATATATATATATATATATATATATATATAT | 17849 | |
| Qy | 3389 | TAGTTTTTTAGTTTTATTAGGTATATCTGGAATGTGATATTTTCAAAATGGGAAGATAT | 3448 | |
| Db | 17848 | ATATTTATATATATATAATAATAAATATATTTATATATTTATATATATATATATAT | 17789 | |
| Qy | 3449 | TGATTTTTHAGACAGACACCTTATAGCATAGCTATCTTATA-ACAGCGCTCAAAACAA | 3507 | |
| Db | 17788 | TATATATATATATATATATATATATAATAATATAATAATATATATATATATATATAC | 17729 | |
| Qy | 3508 | GGTTGGTTGGCTTTATGAACATCTCCGTAAATACCACTACAATATATAGTTTCCAATTC | 3567 | |
| Db | 17728 | ATTATATATATATATATACATTTATATATATATATATATATATATATATATATAT | 17669 | |
| Qy | 3568 | CGTTAATCTTTGCATTATAAAAAATAAATGCAACAATTTTTTTCTTGCTCTGCTT | 3627 | |
| Db | 17668 | ATTATATATATATATATATATATATTTTTTATATATATATATATATATATATAT | 17609 | |
| Qy | 3628 | TTATACCGATCTATTTAAGTGGATCGAAGTTGGTAGTTTATCGTCGCAATATTAATA | 3687 | |
| Db | 17608 | TATATATATATATATATATATATATATATATATATATATATATATATATATATATA | 17553 | |
| Qy | 3688 | TATGCTTTGTTATCGAGATATATAGGTGGAAAAATTTTGCTTTGGATAAAAAGCTAAATAGTAA | 3747 | |

RESULT 12

```

US-09-489-039A-1006
; Sequence 1006, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND
; TITLE OF INVENTION: PNEUMONIAE FOR DI
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,7
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1006
; LENGTH: 1521
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1006

```

| | | | | |
|---------------------------|--------|--------------------|------------|--------------|
| Query Match | 1.9% | Score 132.2; | DB 4; | Length 1521; |
| Best Local Similarity | 54.1%; | Pred. No. 8.5e-15; | | |
| Matches 340; Conservative | 0; | Mismatches 278; | Indels 11; | Gaps 3; |

| | | | |
|------|----|------------------------------------------------------------------|------|
| 746 | Qy | AAATTTTATAAGCCTAGTCATATCATATGATGAAACGACTTTTTCGATATATCTCGAGCGGT | 805 |
| | | | |
| 897 | Db | AAATTTAGCTAAATGGTCATCGAAGATTCCTAAAGCGTACATTTTGATATTGTGGGGCTCT | 956 |
| | | | |
| 806 | Qy | AGTCGGGTAAATTTATTTGTGGTATAGTTTCTATTTTGTAGTTTCCAATTTTCTGTAGAGA | 865 |
| | | | |
| 957 | Db | GGGAATAATTCCTCTTTTCGGCCATTTAAATTTATATTAGTAGAAGAGTAAGAAAGGA | 1016 |
| | | | |
| 866 | Qy | TGCTGGACCGGCTATTTTTCCTCAGAAAAAGAGTTGGACAGAATGGACGCGCATATTTTACATT | 925 |
| | | | |
| 1017 | Db | CGCGGTACAGCTATATACGGGCATGAGCGTGTAGGTAAGGGTGTAAACCAATTTAAATG | 1076 |
| | | | |
| 926 | Qy | CTACAGATTTTCGATTCGATGTATGTTGATGCTAGGAGCGCAAAAAGACATTGCTC | 980 |
| | | | |
| 1077 | Db | CCTCAAAATTCGGTTCAATGGTGTGTAATTCCTAAGAAGTATTAGAAGAGCTGTGGCANA | 1136 |
| | | | |
| 981 | Qy | --AGCCAAAACACAGATGCAAGGGTGGGTATGTTTTTAAATGGCAAAAACGATCCTAGAAT | 1038 |

Db 1137 TGATCCAGAAAGTAAAGCAGAAATGGGATGCAACTTTTCAAACTGAAAAACGATCCTCGCAT 1196
Qy 1039 TACTCCAAATTCGACATTTTCATACGCAAAAAAAGTTTACAGCAGTTTACCAAGTTTATATA 1098
Db 1197 TACTAGTATTTGGTCATTTTTCACGTA-GGACAAGTTTGGATGAACCTTCTCAATTTGTCA 1255
Qy 1099 ATGTTTTAAATGGCGATATAGCTAGTTGGTATACAGTCCACCTACAGTTGATGAATTTG 1158
Db 1256 ATGGCTTTAAAGGGGAGATGATTTAGTAGGGCCACGCGCGATCATTTACAGCGGAGTTAG 1315
Qy 1159 AABAATATCTCTCGGTCAAAGAGACGATTTTAAACCGAGGATTTACAGTCTCT 1218
Db 1316 AGCGATATAACGATGAGCTCGACTACTATCTATTAAAGTAAACCTGGGATGACTGGGCTAT 1375
Qy 1219 GGCAGGTTAGTGGTCGTAGTAAATACACAGCTTCGACGAGTGTAGTTCCGTTGGACTTAG 1278
Db 1376 GGCAGGTCAGCGGGGTAGTGTATGTC---GATTATGATCTCGTGTGTACTTGGATGCGT 1432
Qy 1279 CATACATTGATTAATTTGGACTATCTGGTCAGATATTAATAATTTTAAAGACAGTGAAG 1338
Db 1433 GGTATGTAATAAACTGGTCTATGTGGAATGATATCGCAATCTTGTTCAAAACCATTAATG 1492
Qy 1339 TTGTTATTTGTCAGAGAGGGAAGTAA 1367
Db 1493 TTGTTATTAATAAAAGATGGCGCTTTATTGA 1521

RESULT 13

US-09-949-016-14164
; Sequence 14164, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASES, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14164
; LENGTH: 18773
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14164

Query Match 1.8%; Score 129.2; DB 4; Length 18773;
Best Local Similarity 45.3%; Pred. No. 4.9e-14;
Matches 595; Conservative 0; Mismatches 708; Indels 11; Gaps 3;
Qy 2090 TTTCTAGACAAAAAAGTATGGTGAACATGATCAATCAAGTAGAGTTTGTAA 2149
Db 17099 TTTTCTTATTTTCTATATATATAATATTTTATATAATAATATATAATAATTTTA 17158
Qy 2150 AGAATTTTACAAGATAATAATTTTATTTATAGAAAAATAGATGATTTGTTGAAAAA 2209
Db 17159 TATAATATTTATATAATAATAATTTATATAATAATTTATATAATTTTCTATATAA 17218
Qy 2210 ATTATGAGTTTCTTAAGCAAACTAATTTTACATCAATAATAATTTTGTGAAGA 2269
Db 17219 ATGGTATATAATTTATATAATTTATATAATAATAATATAGATATCTAATAATGATA 17278
Qy 2270 TTAACCAATAGTTGAAAAATTTAAAGAGATCAAGAAATGAATATAAAAAAGATGC 2329
Db 17279 ATATATAACATATAAAAAATATTTATTTATATAATAATTTTATATAATAATTTTAT 17338

Qy 2330 ATATTCGATTAATGCGCTTATCAATAATTTTCTCAGATTTTACTGAGAGGGATACAGATAT 2389
Db 17339 ATATAATATAATATAATATTTTATATAAATTTATTAATATAATAATATAATAATATAAT 17398
Qy 2390 TATCATCTTCTCAGGAGATGCACACCAATTTAGTTCCTTCAGAAATCCTCTATAATTTAT 2449
Db 17399 ATTTTATACATAATTTATTTATATAATAATTAATAATATAATATATC---TTTATCATPA 17455
Qy 2450 TTTAAATATTTCTCAGGATTTTATATGTTGAAATTTTACAAAAGATGAGCAAAAAATAAGAA 2509
Db 17456 TTATCAATTTATATAATAATTAATAATAATAATTTTATATAATAATAATAATTTTATA 17515
Qy 2510 AATAGGATATATGAACAGTTTAAATGTTACAGATTTTCTTAATATATACAGAAAAACT 2569
Db 17516 TATATTTATATTTATATAATAATAATAATGTTTCTATATAATTTTATATAATAATA 17575
Qy 2570 ATTGATATGTTACTGTTTGAATTTTATTAAGAAATGTTACGAGCTTTTGAATACTATTATA 2629
Db 17576 GTATATAATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 17635
Qy 2630 CAAAAGATTTGTTTATTTGATAGAAATAAAAAACATGTTCTAAGAAATTAAGATTTGGTTCTA 2689
Db 17636 TAAAAATATAATAATAATAATAATAATAATAATAATAATAATAAT--GTATAATAATA 17693
Qy 2690 ATTGGGTTTCGCTTCCACATGATTTTGTGGCAATTTCTTTATCAAAATGAAAAACGAAACAG 2749
Db 17694 ATATAATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 17753
Qy 2750 CTTATTTTATTAAGTAACTTAAATGTCAGATGAACTTTTATATACAGCAATTTATAGAA 2809
Db 17754 ATTAATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 17813
Qy 2810 AATATGAATTTTCAAAATAGATTTCTAAATATGAAAAATTTAAGATATATAAAGTGGAAAA 2869
Db 17814 ATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 17873
Qy 2870 AATCAACATCTTCTCCTATTTGCTTTACAGATGATTTCTTATGATGAATGCTAAATGCA 2929
Db 17874 TATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 17933
Qy 2930 GAAATTTTAGGTTTATTTTATTTGCTAGAAAGTTTAAATAAGAAAAATAAATCTAAATTTAAAG 2989
Db 17934 TATATTTTATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 17993
Qy 2990 AAATTTTACTAAAAATAAAATAGTTGATTTTCTGAGAGTAAATGTTATTTAA----- 3043
Db 17994 ATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 18053
Qy 3044 ATATTTTAAATATGACCCGGAATTTTATTTTAAAGTACTTCTGGTTGATTTATTTTAT 3103
Db 18054 ATATAATTTTATATACAAATATGATATAATAATAATAATAATAATAATAATAATAATA 18113
Qy 3104 TCCAGACAAAAGTATGTTATTTTATTAATTTTATTAATTTTATTAATTTTATTTTATATA 3163
Db 18114 TATAATTTTATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 18173
Qy 3164 ATTTTTCGAACTAAAGCTAATAATTTTAAATAATTTTATTTTATTTTATTTTATGTCAT 3223
Db 18174 AATCAATTTATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 18233
Qy 3224 ATATGTTTGTGTTTCTAGTAGTCACAGTATGTTTGTGAAATAAATTTTGAAGATTTAT 3283
Db 18234 ATAGTATTTCTATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 18293
Qy 3284 TGCAGATTTTATGCTCCCATTAATTTGGAATTTTGGCAATTAATGTTATTAATAATTTGTTATTC 3343
Db 18294 TATATAATTTTATAGAAATATATATAATAATAATAATAATAATAATAATAATAATAATA 18353
Qy 3344 ATTTTAAATATTTGATTTTAAATAATTTAAATAATAGTATCTTTTATTTTATTTTAT 3397
Db 18354 TATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATACTT 18407

RESULT 14
US-09-583-110-893
; Sequence 893, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583.110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 893
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-893

Query Match 1.8%; Score 127.4; DB 4; Length 987;
Best Local Similarity 51.2%; Pred. No. 5.7e-14;
Matches 327; Conservative 0; Mismatches 306; Indels 6; Gaps 1;
Qy 6173 AATTCTTAATTTGACCTATATATATAGTAAATATTTTCTAAATGTATAGATAG 6232
Db 15 AATTAGTGTATGTCGCCAGTCTATATGTCGCCAGTCTAGAAAAATCGATGCTTC 74
Qy 6233 CATTGTAATCAGACCTACAAACATATAGAGATTTCTTGGTGAATGACGGTAGTCGA 6292
Db 75 CATTGTAATCAGACCTATCAAAATCTGAAATTTCTTTGATGATGATGTCGACAGA 134
Qy 6293 TAATTCGGAAGAAATTTGTTAGCATATCGAAGAGATAGTCGCATTCGTTATTTAA 6352
Db 135 TGAAGTGTGCTGTTGTGATTCATTCGCTGAAACAGATGACAGGGTGTGCTGCTTCA 194
Qy 6353 AAAAGAGAAGCGGGCTATCAGATGCCGCTTAATTTAGGCAATAGTCGCCCAAGGGTGA 6412
Db 195 TAAAGAGACGAGGATTTGTCGACGACGAAATGATGGATGAGCAGCTCACGGGA 254
Qy 6413 CTACTAGCTTTTATAGACTCAGATGATTTATTTTATTCATTCGAGGTTCATCAAGCTTTACA 6472
Db 255 TTATCTGATTTTATTCAGTCTAGATGATTTATCCATCCAGAAATGATTCAGAGCTTATA 314
Qy 6473 CGAAGCAATGAGAGAGAGATGCCCTTGTGGCAGTTGCTGTTATGATAGGGTAGATGC 6532
Db 315 TGAGCAATTTTCAAGAAGATGC-----GGATGTTTCAGCTGTGGTGTGATGATGT 368
Qy 6533 TTCGGGGCATTTCTTAACAGCAGACGCGCTTCTTACAAATCAGGCTGTTCTGAGCGGCAG 6592
Db 369 CTATGCTAATGATGAAAGCCACAGTCAGCCAAATCAGGATGACTATTTTCTGTGATTC 428
Qy 6593 GAATGTTTGAAGAGCTGTAGAGCGGATGTCATCGCTTGTGTGGCTGCTGATATAA 6652
Db 429 TCAACATTTCTAAGGAATACCTCATAGGTGAAAAAATACCTGGGACGATTTGCAATAA 488
Qy 6653 ACTCTATAAAAAAGAACTATTTGAAGATTTTCGATTTGAAAGGGTAAGATTCATGAAGA 6712
Db 489 GCTAATCAGAGACAGATTCGAATCTGACCTATCCTTCTTCAAGGGGTGATTTACGAAGA 548
Qy 6713 TGAATACCTTCACTTATCGCTTGTCTATGATTTAGAAAAAGTTGCAATAGTTAAGGAGTG 6772
Db 549 TGCCTATTACATTTTATTAATCAAGTTGGCCCAAGAGTATGTGCTTAATACTAAACC 608
Qy 6773 CTGTACTATTATGTTGACCGAGAAAAATAGTATACAAAC 6811
Db 609 CTATTATTACTATTTTCATAGAGGGGATAGTATTACGAC 647

RESULT 15
US-09-107-433-2107
; Sequence 2107, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/107,433
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 893
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-107-433-2107

Query Match 1.8%; Score 127.4; DB 4; Length 996;
Best Local Similarity 51.2%; Pred. No. 5.8e-14;
Matches 327; Conservative 0; Mismatches 306; Indels 6; Gaps 1;
Qy 6173 AATTCTTAATTTGACCTATATATATAGTAAATATTTTCTAAATGTATAGATAG 6232
Db 24 AATTAGTGTATGTCGCCAGTCTATATGTCGCCAGTACCTAGAAAAATCGATGCTTC 83
Qy 6233 CATTGTAATCAGACCTACAAACATATAGAGATTTCTTGGTGAATGACGGTAGTCGA 6292
Db 84 CATTGTAATCAGACCTATCAAAATCTGAAATTTATTTCTTGTGATGATGTCGTCACAGA 143
Qy 6293 TAATTCGGAAGAAATTTGTTAGCATATCGAAGAGATAGTCGCATTCGTTATTTAA 6352
Db 144 TGAAGTGTGCTGTTGTGATTCATTCGCTGAACAGAGTGAACAGGGTGTGCTGCTTCA 203
Qy 6353 AAAAGAGAAGCGGGCTATCAGATGCCGCTTAATTTATGTCATAGTCGCCCAAGGGTGA 6412
Db 204 TAAAGAGAAGAGGATTTGTCGCAAGCAGAAATGATGGATGAAGCAGGCTCACGGGA 263


```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: EP 95201669.9
  FILING DATE: 20-JUN-1995
  ATTORNEY/AGENT INFORMATION:
    NAME: Fanucci A., Allan
  REGISTRATION NUMBER: 30256
  INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 14602 base pairs
      TYPE: nucleic acid
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 352..1803
    OTHER INFORMATION: /product= "epsa"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 1807..2535
    OTHER INFORMATION: /product= "epsb"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 2547..3239
    OTHER INFORMATION: /product= "epsc"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 3249..3995
    OTHER INFORMATION: /product= "epsd"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 4051..4731
    OTHER INFORMATION: /product= "epsf"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 4898..5854
    OTHER INFORMATION: /product= "epsf"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 6425..7540
    OTHER INFORMATION: /product= "epsg"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 7736..8212
    OTHER INFORMATION: /product= "epsh"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 8221..9192
    OTHER INFORMATION: /product= "epsi"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 9285..10364
    OTHER INFORMATION: /product= "epsj"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 10392..11339
    OTHER INFORMATION: /product= "epsk"
  FEATURE:
    NAME/KEY: misc_feature
    LOCATION: 11302..12222
    OTHER INFORMATION: /product= "CDS (eps l) covering CDS"
  OTHER INFORMATION: (eps k) on nucleotides 10392-11339"
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 12233..13651
    OTHER INFORMATION: /product= "epsm"
```

```
NAME/KEY: misc_feature
LOCATION: 13732..14305
OTHER INFORMATION: /function= "CDS on the
OTHER INFORMATION: complementary strand"
OTHER INFORMATION: /product= "orfz"
FEATURE:
NAME/KEY: terminator
LOCATION: 230..252
FEATURE:
NAME/KEY: promoter
LOCATION: 274..302
FEATURE:
NAME/KEY: RBS
LOCATION: 340..345
US-08-597-236-1

Query Match 1.7%; Score 122.2; DB 1; Length 14602;
Best Local Similarity 48.5%; Pred No. 8.6e-13;
Matches 337; Conservative 0; Mismatches 358; Indels 0; Gaps 0;

Qy 6118 TTTTAAAAAATATTATGGTTATAATAGGAAGATATCATGATACACTATGTAATAAATTT 6177
Db 8183 TTATCTAATGGAGAAATTAGAAAGCGTGGCTAATTAATAATGTAATCTTAAAGTCTAATCT 8242
Qy 6178 CTATAATTGTACCTATATATAATAGTAAATAATTTATCTAAATGTATAGATAGCATTTG 6237
Db 8243 CTATTGTTATCCAGTATATAATGTAGAGAAATATTTAGAAAAATGTTTGCATCTGTTTC 8302
Qy 6238 TAAATCAGACTACAAACATATAGAGATTCTTCGTTGAATGACGCTAGTAGCGATTAATT 6297
Db 8303 AAATCAGACTTAAATATTTTGAAGTGAATTTTAGTGAATGATGGCTCAACCCATTCAT 8362
Qy 6298 CGGAAGAAATTTGTTTAGCATATCGGAAGAAAGATAGTCGCATTTCGTTATTTTAAAAAAG 6357
Db 8363 CACTTTCATATGCGAAAAATTTGTTAATCAGGATAAAGATTTTCTGTTTTTCTAAAG 8422
Qy 6358 AGAACGGCGGCTATCAGATGCGCGTAATATGATGATAGTTCGCGCAAGGGTACTACT 6417
Db 8423 AAATGGTGTATGTCATCTGCACGAAATTTTGAATTTAAAAAGGCTAAAGGATCGTTTA 8482
Qy 6418 TAGCTTTTATAGACTCAGATGATTTTATTCATTCGGAGTTTCATCCAAAGTTTACACGAAG 6477
Db 8483 TCACATTTGTAGATAGTGTAGTACATAGTAAAGATTTATCTTCTCATTTGGTAGCTG 8542
Qy 6478 CAATTGAGAGAGAAATGCCCTTGTGTCAGTTGCTGTTATGATAGGTAGATGCTTCGG 6537
Db 8543 GGATAAAAGTGAGACCTCTATAGTTTGTTCAAAGTTTTTCTTGTAGATGAAAAAGGAA 8602
Qy 6538 GGCAATTTCTTAACAGCAGAGCGGCTTCCTACAAATCAGGCTGTTTCTGAGCGGCAAGATG 6597
Db 8603 GTTTATTGACTAAAAAAGAGGACCTTAAAAAGAAATCAGAAATCGTTTCAATTGAGGAAA 8662
Qy 6598 TTTGTAAGAGCTGTAGAGCGGATGGTCATCGCTTTGTGGTGGCTGTATTAATAACTCT 6657
Db 8663 GTATTAAAAATTTCTCTGTTGCAACAAAATGGCTATGATCTGCTGTCTGGGGAAAAATTA 8722
Qy 6658 ATAAAAAAGAACTATTTTGAAGATTTTCGATTTGAAAAAGGTAAGATTCATCAAGATGAAT 6717
Db 8723 ACCCGTTTCTTTCTTTTGAACAAATTTCTTCCAGAGGAAACTTTTACGAGATATGG 8782
Qy 6718 ACTTCATCTATCGCTTCTCTATGAGTTAGAAAAAGTTGCAATAGTTAAGGAGTGTCTGT 6777
Db 8783 GAACAACCTTACAAATTACTTAAATTTGCAAGTGAAGTGGTCTTCTTGTGATCGGTATGATT 8842
Qy 6778 ACTATTATGTTGACCGAGAAAAATAGTATCAAACT 6812
Db 8843 ATGCCTACGTACAGCGACCTTAATAGTATCATGAAT 8877

RESULT 20
US-08-746-682A-1
; Sequence 1, Application US/08746682A
; Patent No. 5786184
```


| | | | |
|----|------|------------------------------------------------------------------|------|
| Qy | 6658 | ATAAAAAGAACTATTATTTGAAGATTTTCGATTTGAAAAAGGTAAAGATTTCATGAAGATGAAT | 6717 |
| Db | 8723 | ACCCGGTTTCTTTCTTTTGAACAATTTCTTTCCAGAAGAAACTTTTACGAAGATATGG | 8782 |
| Qy | 6718 | ACTTCACCTTATCGGTTTCTCTATGAGTTAGAAAAAGTTTGCANATAGTTTAAAGAGTGCTTGT | 6777 |
| Db | 8783 | GAACACACTTCAAAATTAATAAAATTGCAAGTGAAGTGGTCTTCTTGGATCGGTATGATT | 8842 |
| Qy | 6778 | ACTATTATGTTGACCGAGAAANATAGTATCAACT | 6812 |
| Db | 8843 | ATGCCTACGTACAGCGCACTTAATAGTATCATGAAT | 8877 |

```

RESULT 21
US-09-495-406-1/c
; Sequence 1, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-0001100S
; CURRENT APPLICATION NUMBER: US/09/495,406
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11474
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
; - OTHER INFORMATION: including LOS biosynthesis locus
US-09-495-406-1

```

| Query Match | 1.7k; | Score | 121.4; | DB | 4; | Length | 11474; |
|-----------------------|--------|-------------------------------------------------------------------|----------|------------|------|--------|--------|
| Best Local Similarity | 54.8k; | Pred. No. | 1.1e-12; | | | | |
| Matches | 324; | Conservative | 0; | Mismatches | 261; | Indels | 6; |
| Gaps | 4; | | | | | | |
| QY | 4998 | TAATGAGTTCCTGCGCAATATTTAGAGTCCTTATCGAAAGAAATACGTAGATATCCATTTAT | 5057 | | | | |
| DB | 6132 | TAAATACATCAAAATCATTTGGTAGCTTGGAAATATCAAAATCTTTTAAACTTGGTCCAT | 6073 | | | | |
| QY | 5058 | TAAAGCGAAAGATATTTATCAAGAAAGCATTTAGTTTACGTGTGATTTGATGAAATTTTC | 5117 | | | | |
| DB | 6072 | TTCCAGCAATATAAAGCTTTTTTTCATGATTATTTTCCATCAAGATTATTTTTTAAACGATTT | 6013 | | | | |
| QY | 5118 | GCCTTAAACTATATGTTAATGTTTATATAGAATTTTCAAAGCAGCTAGAGGTAAATAATCGAT | 5177 | | | | |
| DB | 6012 | TATCAANATAGATTTATGTTTATA-ATTAAATTTATATAAATAAAGAAATAAATATGTTTT | 5954 | | | | |
| QY | 5178 | AAAATTAGTGTTATTTGTTCCAGTTTATAATGTAGATAAAATATTTTAAAGTAGTCTGTATAGAA | 5237 | | | | |
| DB | 5953 | AAAAATTCATCATCTTACCAACTTATATGTGGAAACAATATATAGCAAGGGCAATAGAA | 5894 | | | | |
| QY | 5238 | AGCATTTTAAATCAAAATTTAATAATATAGAATATTTATTTGATAGATGATGCTCTGTATA | 5297 | | | | |
| DB | 5893 | AGCTGTATCAATCAGACTTTTTTAAAGATATAGAATAATTTGATGTGATGATTTGTGGAAAT | 5834 | | | | |
| QY | 5298 | GATGATTCTGCTAAAAATATGCAAGGAATATGCGAAAAAGATAAAAAGGTAAAAATTTTTT | 5357 | | | | |
| DB | 5833 | GATAATAGTATATAATATAGCCAAAGAATACTCTAAAAAAGACAAAAAATAAATAATC | 5774 | | | | |
| QY | 5358 | TTTCACT---AATCATAGTGGAGTATCAAAATGCTAGAAATCATGGATAAAGCCGAGTACA | 5414 | | | | |
| DB | 5773 | CACAATGAAAAAAACCTTAGTCTTTTTAAGAGCAGAGATATGAAGGTGTGAAGGTAGCAAAC | 5714 | | | | |
| QY | 5415 | GCTGAATATATTTATGTTTGTGTGACTCTGATGATGTTGTTTTCATAGTAGATATGATAGAAAAA | 5474 | | | | |

```

Db      5713  TCTCCTTATATAATGTTTTTAGATCCTGATGATTATTATGGA-ACATAATGCTTTGTGAAGA 5655
Qy      5475  TTATATTTTAAATATATAAAAGTAGAAGTGATTTATCTGGT-TGTTTGTACGCTACTTT 5533
Db      5654  GTCTATAAAAAATTTTAGATCAACAGGATGAAGTTGATTAGTGTGTTTTTCAATGCTATTGT 5595
Qy      5534  TTCAGAAAATATAAAATAATTTTGAAGTGGAATAATCCAAATATTGATTTTGA 5584
Db      5594  TGAAGTAATGTTATTTCATATATAAAAAGTTTGACTTTAATTCTGGTTTTTA 5544

RESULT 22
US-09-816-028A-1/c
; Sequence 1, Application US/09816028A
; Patent No. 6693705
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11474
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
; OTHER INFORMATION: including LOS biosynthesis locus
; US-09-816-028A-1

```

| Query Match | 1.7%; | Score 121.4; | DB 4; | Length 11474; |
|-----------------------|-----------------|-----------------------------------------------------------------|-----------|---------------|
| Best Local Similarity | 54.8%; | Pred. No. 1.1e-12; | | |
| Matches 324; | Conservative 0; | Mismatches 261; | Indels 6; | Gaps 4; |
| QY | 4998 | TAATGAGTCTCTGCCAATATTAGAGTCCTTATCGAAAGAAATACGTAGATATCCATTTAT | 5057 | |
| DB | 6132 | TAATATACATAAATCATTTGGTAGTCTTGAATATCAATTTCTTTTAAACTTGGTCCAT | 6073 | |
| QY | 5058 | TAAAGCGAAAGANATTTTATCAAGAAAGCATTTAGTTCAGTTGTGTATTGTGATGAAATTTTC | 5117 | |
| DB | 6072 | TTCCAGCAATAATAACTTTTTTCATGATTATTTTCCATCAAGATTTATTTTAAACGATTT | 6013 | |
| QY | 5118 | GCTTAACTATATGTAACTGTATATAGAATAATTCAAAAGCAGTAGAGGTAAATAATGGAT | 5177 | |
| DB | 6012 | TATCAAAATAGATTTTATGTGTATA-ATTAAATTTATATAAATAAAGGAATAAATATGTGTT | 5954 | |
| QY | 5178 | AAAAATTAGTGTTATTGTTCCAGTTTATAATGTAGATAAAATATTTAAAGTAGTTGTATAGAA | 5237 | |
| DB | 5953 | AAAAATTTCAATCATCTTACCACCTTATAATGTGGAACAATATATACAAAGGCAATAGAA | 5894 | |
| QY | 5238 | AGCATTTATTAATCAAAATTTATAAANATATAGAAATATTATTTGATAGATGCTCTGTGA | 5297 | |
| DB | 5893 | AGCTGTPATCAATCAGACTTTTAAAGATATAGAAATAAATTTGTAGTTGATGATTTGGGAAAT | 5834 | |
| QY | 5298 | GATGATTTCTGCTAAAATATGCAAGGAATATGCAGAAAAAGATAAAAGAGGTAAATAATTTTT | 5357 | |
| DB | 5833 | GATAATAGTATAAATATATACCCAAAGAAATCTCTAAAAAAAGACAAAAGATAAAAAATAATC | 5774 | |
| QY | 5358 | TTCACT---AATCATAGTGGAGTATCAAAATGCTAGAAATCATGGAATTAAGCGGAGTACA | 5414 | |
| DB | 5773 | CACAATTGAAAAAAAACCCTTAGTCTTTTAAAGACAAGATATGAAGGTGTGAAGAGTAGCAAC | 5714 | |
| QY | 5415 | GCTGAATATATTATGTTTGTGACTCTGATGATGTTGTTGATAGTAGATTTAGTAGAAAAA | 5474 | |

Db 5713 TCCTCCATATAATGTTTGTAGTCTCTGATGATTAATTGGA-CTAAATGCTTGTGAGA 5655
Qy 5475 TTATATTTTAAATATAAAGAGTAGAAGTATTAATCTCGT-TGTTTGTACGCTACTTT 5533
Db 5654 GTGTATAAAAAATTTAGATGAACAGAGTGAAGTTGATTTAGTGTGTTTTCATGCTATTGT 5595
Qy 5534 TTCAGAAATATAAATTTTGAAGTGAATAATCCAAATATTCATTTTGA 5584
Db 5594 TGAAGAGTAATGTTATTTTCATATAAAAAAGTTTGACTTTAAATTCCTGGTTTTTA 5544

RESULT 23

US-10-303-162-1/c
; Sequence 1, Application US/10303162
; Patent No. 6723545
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; FILE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,162
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11474
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
; OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-162-1

Query Match 1.7%; Score 121.4; DB 4; Length 11474;
Best Local Similarity 54.8%; Pred. No. 1.1e-12;
Matches 324; Conservative 0; Mismatches 261; Indels 6; Gaps 4;
Qy 4998 TAATGAGTTCCTGCCAATATTAGAGTCTTATCGAAAAGAAATAGTAGATATCCATTTAT 5057
Db 6132 TAAATACATCAAAATCAATTTGGTAGTCTTGAATAATCAATTTCTTTTAAACTTGGTCCAT 6073
Qy 5058 TAAAGCGAAAGATATTTATCAAGAAAGCATTTAGTTAGTGTATTTGTATGTAAGAAATTTTC 5117
Db 6072 TTCCAGCAATAATAACTTTTTCATGATTAATTTTCCATCAAGATTTATTTTAAACGATTT 6013
Qy 5118 GCCTAAACTATATGTAATGTTATTAAGAAATTTTCAAAAGCAGTAGAGGTAAATCGAT 5177
Db 6012 TATCAAAATAGATTTTATGTTTATA-ATTAAATTAATAAAATAGGAATAAATATGTTT 5954
Qy 5178 AAAATTAGTGTATTGTTCCAGTTTATAATGATAGATAAATATTTTAACTAGTTGATAGAA 5237
Db 5953 AAAATTTCAATCATCTTACCAACTTTATAATGCGAACATATATAGCAAGGCAATAGAA 5894
Qy 5238 AGCATTTAATCAAAATATATAAAATATAGAAATATTTTGTATGATGATGCTCTGTA 5297
Db 5893 AGCTGTATCAATCAGACTTTTAAAGATATAGAAATAATTTGTAGTTGATGTTGGAAT 5834
Qy 5298 GATGATTCGCTAAATATGCAAGGAATATCGAGAAAGATATAAGAGTAAATTTTTT 5357
Db 5833 GATTAATAGTATAATATAGCCAAAGATATCTTAAAGAAAGCAAAAGATAAATAATATC 5774
Qy 5358 TTCACT---AATCATAGTGGAGTATCAAAATGCTAGAAATCATGGAATAAAGCGGAGTACA 5414
Db 5773 CACAATGAAAAAAAATCTTAGGTCCTTTAAGACCAAGATATGAAGGTGTGAAGTAGCAAC 5714

Qy 5415 GCTGAATATATATGTTTGTGACTCTGATGATGTTGTTGATAGTAGATAGTAAAAA 5474
Db 5713 TCCTCCATATAATGTTTGTAGTCTCTGATGATTAATTGGA-CTAAATGCTTGTGAGA 5655
Qy 5475 TTATATTTTAAATATAAAGAGTAGAAGTATTAATCTCGT-TGTTTGTACGCTACTTT 5533
Db 5654 GTGTATAAAAAATTTAGATGAACAGAGTGAAGTTGATTTAGTGTGTTTTCATGCTATTGT 5595
Qy 5534 TTCAGAAATATAAATTTTGAAGTGAATAATCCAAATATTCATTTTGA 5584
Db 5594 TGAAGAGTAATGTTATTTTCATATAAAAAAGTTTGACTTTAAATTCCTGGTTTTTA 5544

RESULT 24

US-10-303-134-1/c
; Sequence 1, Application US/10303134
; Patent No. 6825019
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; FILE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,134
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 11474
; TYPE: DNA
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: 11.5 kb PCR product from C. jejuni OH4384
; OTHER INFORMATION: including LOS biosynthesis locus
US-10-303-134-1

Query Match 1.7%; Score 121.4; DB 4; Length 11474;
Best Local Similarity 54.8%; Pred. No. 1.1e-12;
Matches 324; Conservative 0; Mismatches 261; Indels 6; Gaps 4;
Qy 4998 TAATGAGTTCCTGCCAATATTAGAGTCTTATCGAAAAGAAATAGTAGATATCCATTTAT 5057
Db 6132 TAAATACATCAAAATCAATTTGGTAGTCTTGAATAATCAATTTCTTTTAAACTTGGTCCAT 6073
Qy 5058 TAAAGCGAAAGATATTTATCAAGAAAGCATTTAGTTAGTGTATTTGTATGTAAGAAATTTTC 5117
Db 6072 TTCCAGCAATAATAACTTTTTCATGATTAATTTTCCATCAAGATTTATTTTAAACGATTT 6013
Qy 5118 GCCTAAACTATATGTAATGTTATTAAGAAATTTTCAAAAGCAGTAGAGGTAAATCGAT 5177
Db 6012 TATCAAAATAGATTTTATGTTTATA-ATTAAATTAATAAAATAGGAATAAATATGTTT 5954
Qy 5178 AAAATTAGTGTATTGTTCCAGTTTATAATGATAGATAAATATTTTAACTAGTTGATAGAA 5237
Db 5953 AAAATTTCAATCATCTTACCAACTTTATAATGCGAACATATATAGCAAGGCAATAGAA 5894
Qy 5238 AGCATTTAATCAAAATATATAAAATATAGAAATATTTTGTATGATGATGCTCTGTA 5297
Db 5893 AGCTGTATCAATCAGACTTTTAAAGATATAGAAATAATTTGTAGTTGATGTTGGAAT 5834
Qy 5298 GATGATTCGCTAAATATGCAAGGAATATCGAGAAAGATATAAGAGTAAATTTTTT 5357
Db 5833 GATAATAGTATAAATATAGCCAAAGATATCTTAAAGAAAGCAAAAGATAAATAATATC 5774
Qy 5358 TTCACT---AATCATAGTGGAGTATCAAAATGCTAGAAATCATGGAATAAAGCGGAGTACA 5414

Db 5773 CACATGAAAAAACTTAGTCTTTTAAAGAGCAAGATATGAAGGTGTGAAAGTAGCAAAAC 5714
Qy 5415 GCTGAATATATTTATGTTTGTGACCTCGATGATGTTGTTGATAGTAGATAGTAGCAAAAA 5474
Db 5713 TCTCTTAATAAAGTTTTTATAGTCTCGATGATTTTGGGA-ACATAAGCTTGTGAAGA 5655
Qy 5475 TTATATTTTAATATATATAAAGTAGAAGTAGATTTATCTGGT-TGTTTTAGCTACTTTT 5533
Db 5654 GTGTATAAAAAATTTTAGATGAACAGGATGAAGTTGATTTAGTGTGTTTTCAATGCTATTGT 5595
Qy 5534 TTCAGAAATATAATAATTTTGAAGTGAATAATCAATATTGATTTTGA 5584
Db 5594 TCAAGATTAATGTTATTTTCATATAAAAAAGTTTGACTTTTAATTTCTGGTTTTTA 5544

RESULT 25

US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Kin-zhan
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13
Query Match 1.7%; Score 119.2; DB 2; Length 19124;
Best Local Similarity 42.4%; Pred. No. 3.2e-12;
Matches 1497; Conservative 0; Mismatches 1983; Indels 48; Gaps 14;
Qy 2833 TCTAATAATGGAATTTTAAAGATATATAAAGTGGAAAAAATCAATCTTCTCTTAATGTC 2892
Db 3504 TGTGAATTTTTCATTTAATATGCTATGATCATTTTGATAATTAATTTTTTTTATAATATT 3563
Qy 2893 TTACAGATGATCTTATGATGAATTCGTAATGCAAGAAATTTAGGTTTTTTTATTGCT 2952

Db 3564 ATATTTTTTATACCTTGGATTTCTTACATTTGTTTTATTTATTTATATGATTTATTTATTTAT 3623
Qy 2953 AGAAAGTTTAAAAATAGAAAAATAAAATCTAAATTTTAAAGAAAAATTTTACTTAAAAATAAAT 3012
Db 3624 ATACTTATATATATATATATATTTTACATTAAGATATATATATATGATCTATCTATCTATC 3683
Qy 3013 AGTTGATTTTGTGAGAGTAATGTATGTTTAAATTTATTTAAATATATGACCGGAATATTTTA 3072
Db 3684 TATCTATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 3743
Qy 3073 TTTTAAAGTACTTCTGGTTGATTTATTTTATTTCCAGAGCAAAAGTATGTTT---TTTTAT 3129
Db 3744 GTGATGATTTAATAATAAACCTTATTGAAGAGAGATAGAAACATAATAATATATATTTAAAT 3803
Qy 3130 TAAATTTTATGAAATTTTAAATTTTATTTTATATAAAAAATTTTGAATACTAAGCTTAAATTA 3189
Db 3804 AGAATTTTCAATTTTATTTGTTATATATATATATAAAAAATAAGAAATTTGAAAAAGTAAATTT 3863
Qy 3190 AAAATGAAATTTTATTTGTTTATTTATGTTCTATATATATATATGTTTGTGTTTTCAGTAGT 3249
Db 3864 ACATGATAATGTTATTTTATTTTATTTTGTGTTTGTGTTTATATATATTTTATAAAAAATGTT 3923
Qy 3250 GTATGTTTGTGAAATATAAATTTTGAAGAATTTTTCGAGATTTTACTGCTCCCATATTTT 3309
Db 3924 ATATAAGTTGTTTATATAAATTTTAAATATGTCACCATTTAGCTTTTCCATTTATACAAATAT 3983
Qy 3310 GGATTTATGCAATATATATATATATATTTGTTATTTTATATATATATATATATATATATAA 3369
Db 3984 ATATTTCTCTCAATTA--GAATCTGAATATTTTATGTTATATAAAAAAAGTATAATAATA 4041
Qy 3370 TAAAAAATAGTATCTTTTTTATGTTTATTTTATTTAGTTATATCTGATTTATATATTA 3429
Db 4042 AAATATCTAAGATTTTCTAATTTGTTTAAATTTTATATAAATTTTAAATTTTATACGATA 4101
Qy 3430 TTCAAAATGGAAGATATGTTATTTTATGACAGA-CACCTTATAGGACTAGACTATCTT 3488
Db 4102 GAATAAATTAATAACAACATATATATATATGTTTATCTTAAAGAACCTATTACAAATATAGT 4161
Qy 3489 ATAACAGCGTCAAAACAAGGTTGGTTGGCTTTATGAATCTTCTACGTTAAATACCCT 3548
Db 4162 AACAACTGGTTCTCTTTTATTAATAACAATAAGAATGTGTAAAGGATAGTTGTTAAA 4221
Qy 3549 ACAATATAGTTTCAATTTCCGTTAATCTTTGCACCTTATAAAAAATAAATAAATGCAACA 3608
Db 4222 GGCTTTTAAATATGTTTAAATGTTTGAAGATATATATATATATATATATATATATATAT 4281
Qy 3609 TTTTCTTGTGCTTGTCTTTTATACCGATCTATTTAAGTGGATCGAGAAATTTGGTAGTTTA 3668
Db 4282 CAACCTTTGCATAATTTGTAATTTAAAAAATAATATATAAAGAAATTTTATAAATAATATT 4341
Qy 3669 TCGGTAGCAATATTAATTTATATGTTTGTATGGAGATATATAGTGGGAAATTTTCTGTTGG 3728
Db 4342 ATAAAAAATTAAGCATAAATCTCAATAAATTTTTTTTTTATTAATTTTAAATTTTATTTTA 4401
Qy 3729 ATAAAAAGCTAATAGTAATATTTGTAATCTACTTATTTATTTTAAATCTGAAATGCTT 3788
Db 4402 TTGTTCTTAAATATATGATTAAGAAATATTA-TTTGTGCTAATATATATTAAGATATT 4460
Qy 3789 TACCATGAAATTTTGGCTGTTTATAAATTTCTAGAGAAATCAAGTAACGAAGCTAGATTATT 3848
Db 4461 TCTAATATTAATTTATATATATATATATTTAAAGATTTTAAAGAAATTTTAAAGAAATTTT 4520
Qy 3849 ATTTATCA-----AGGAAGTATTGATTAAGTATTAGAAAAACAATATTTTATT 3896
Db 4521 TATTATAATGAATAATGATGATGAGTATATATAAATTTGATGACAAAAAATAAATTTT 4580
Qy 3897 GGATATGGAATATCGGATATTTTCAAGTACGGAATTTGGCTCGGAAGTCAATTCAGGCTAT 3956
Db 4581 TAAATGGAATAATATGCATATAAATAAATACTATATAGTATAAATTTGGTGAATAGTTGTA 4640
Qy 3957 ATATCAATTTTTTATATAAATCAGGAATAGTTGGGTTGATTTTCTAGTATGTTTCTTTT 4016

Db 4641 ACTTATACAAACATGTTGCAATTCATATAATTTAGAGATTATGTAAATTTGTTTATGATCGT 4700
Qy 4017 TATGTTTATAAAAAGTTATGGAGTTTAAATCGGGAAACAGCAGCTATTTTATTTTACATCA 4076
Db 4701 AATATATATTAATAATAATTTGTTTTTATGATGTATGTTATCTTAATAATATATTCATAT 4760
Qy 4077 TTAGCCATATTTTTCATATATGAACAATAGATCCGATTATTTATATATTAGTACTATTC 4136
Db 4761 GTAGTCATAGTGTCAATGAATATAAAAATUGGTATATTTTATATTATGTTATATTAAT 4820
Qy 4137 TTTTCTTCAATPAGGTATTTGGAAATAATATATAAATTTTAAAAAGGATATCGAGACAAAAAT 4196
Db 4821 AAGTAACACAGCAATATATATAGTAATAAATAGAGAAATATATATTTTATGTTAT 4880
Qy 4197 GAATGATTTAATTTCAAGTATTTGACCAATTTATTAATGTCCAGATTAATCTTGATATAATG 4256
Db 4881 ATATTATTAGTTATATAAAGGGGAAATTCATATAATTTATGAAAAATTTTGTATATGA 4940
Qy 4257 TATTAAACAGTATTATTAACCAACATATACTAATTTTAGAGGTTTATCTCGTAAATGATGG 4316
Db 4941 TATAGTTATAGTTTAAAAAATAAATAAATAAAGCAAAATGGAAGCAATAAAAAATGT 5000
Qy 4317 AAGTACTGATGATTTCTGAGAAAAATTTGCTTAAACTATATGAAGAACGATCGAAGAAATTA 4376
Db 5001 TACTGTAATAGGATAAATAATATATATATAAATGTTTATTTTATCTTAAAAAGTTCCTA 5060
Qy 4377 ATATTACAAGAAATTAATGGCGGTCTGACGAGATGCTCGAAATTTCCGAGCTGAGAACATGC 4436
Db 5061 TTATAACATTTAAAAAATAATTTGTCCCATTTTATATAAATAATTAACATACATTTACATAATGA 5120
Qy 4437 AACAGGTAAATATATGCTTTTGTGCGATCTCGATG-----ACTATATAGAGTTGCAATG 4491
Db 5121 AATTCGATTTTGTGTTTTTGTGATGAATATATGGAATTAATTTATATGTTGAATGCG 5180
Qy 4492 TTCGAGAAATGATGATAATATAAATGAGTATTAATGCCATATAGCAGAGATAGATTTT 4551
Db 5181 TTCTATATAATAATAATTTTATTTAAAAAATGAATAATGAATAAATATCCCTGA 5240
Qy 4552 TGTTTATGACGAAAAACGGGTATACAAAGAAAAAGAAATAGTAATTTTCAATGCTTA 4611
Db 5241 TTTTGTAGTCCAAATAGCTTAATAATAATATGAGTCTCATATATATATATATATATCTTT 5300
Qy 4612 ACAGAGAGAGAGCTGTAAGAATTTTGTGCAGGATC-----TAAATAGA 4658
Db 5301 ACAACAAGTATAAGTAATATTTTATTTAAATCTTAATAAGAAAAATAAATAAATAAAT 5360
Qy 4659 AAATAAATGTTTGTGCAAGCTTTTATCCAGAGATTTATATAAGATATAAATAATCCAAAT 4718
Db 5361 AAGAATCTGAATAATAGTCATATTATACATTTTAAAAATGTAACATAAATTACAAAT 5420
Qy 4719 TAATAATAGAGTATTTGGTGAAGATTTGCTTTTAAATTT--GGAGGCTTTGAACAATGTA 4776
Db 5421 ACGTAAACATGTATTATAGAAATAATAAGAAATTAATAATTAAGGATAAATAATAATTTA 5480
Qy 4777 ACAGCTGTAGTTGATACCTAGAGATATTTATTAATTTATGTCATTCGTACAGTTGCG 4836
Db 5481 AAATTAATTTTTTTATGTCAAATTTATGTTATATATATATATAATTAACATGATTAGTTTT 5540
Qy 4837 CTATTAAATCAGAAATTTCTTATATAATAATTCATTTTAGTTCACAGATTGGAAATTAAC 4896
Db 5541 TTGAAAAATTTTAAATATCATATAATAATAATAATAATTTAGTTAAATAATAGTATTTTCAAT 5600
Qy 4897 CCCTTTAAGTTAAAAAGAGGTTTATGTCATTTATTTTGAATGCAAAAGTTTATTAAGAGAAAG 4956
Db 5601 ACAAATACTAACTTATAAGATATATCATATAAATATATATATATATATATATTTATGTTT 5660
Qy 4957 GTTAAATGTTTAAACAAAATGTTTCAACAGATTTGTTTGGATATGATGTTCTCCCAATA 5016
Db 5661 TTGATGGGTATATATAGGCTTAATAGGTATATATATGTTGTTTCAATATATATATTTATATGT 5720
Qy 5017 TTAGAGTCTTATCGAAAGAAATACGTAGATATCCATTTTATTAAGCGCAAAAGATATTTA 5076
Db 5721 GAATAGATACATATAAGTT--AATATATTTATTTGTTGTTATGTTCTGTGTTAAGATAGATA 5779

Qy 5077 T-CAAGAAAGCATTTAGTTACGTTGTATTTTGATGAAATTTTCGCCTAAACCTATATGTAAT 5135
Db 5780 TGCATTTACAGTTAAGGTTATAGTTTTTTTTTTTTTTTTTTTTTTTGTACATATATAAAAAAT 5839
Qy 5136 GTTATATAAGAAATTTCAAAAGCAGTAGAGGTAAAAATGATAAAAATTTAGTGTATTGTT 5195
Db 5840 AGATACTAACATATGCTATTTTACAAGATAATATTTTGTATAAATAATATATATATATA 5899
Qy 5196 CCAGTTTATTAATGTAGATAAATATTTAAAGTGTGTATAGAAAGCAATTTAATAATCAAAAT 5255
Db 5900 TATATATAAA--GACATTAAAACTATACTAATAAGGTAATTAGTTTTTATTATATCATCCT 5956
Qy 5256 TATAAAAAATATAGAAATTTTATGATAGATGATGGCTCTGTAGATGATTTCTGCTAAAAATA 5315
Db 5957 TTTATTATTATAAATTTTTTTTGTGTTTACTCTTCTGTCGTTCTTTTTTGTGTTATTATAATA 6016
Qy 5316 TGCAGGAATATCGCAAAAAAGATAAAAAGTAGTAAAAATTTTTTTCACATAATCATAGTGA 5375
Db 6017 ACAATATATAACAATATCTAGTATTTTGGAAATATAAATAAATTTATTTCTACATATATGCAT 6076
Qy 5376 GTATCAAAATGCTAGAAATCATGGAATAAGCGGAGTACAGCTGAATATATATTGTTTGT 5435
Db 6077 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 6135
Qy 5436 GACTCTGATGATGTTGTTGATGATGATAGTAGTAGTAAAGAAATATATTTTTTAATATATATAA 5495
Db 6136 TACTATTTTATATACATGCAATTTTTTATATATATTTTATATATATTTTAAAGATATATTTAAT 6195
Qy 5496 AGTAGAAGTGAATTTATCTGTTGTTGTACGCTACITTTTTTTCAGAAAAATATAAATAATTTT 5555
Db 6196 ATTTATATAGTACATATATGTTATTTATATATAAACAATTTTTCATTTTATATATAAATAT 6255
Qy 5556 GAAAGTGAATAATCAAAATTTTGAAGCAATTAATACCGTGCGAGGACATGGGAGAA 5615
Db 6256 ATAGAACATGAACATTTTATTAATAACTCATATTTGAAATATATATATTTAATAATGCTAT 6315
Qy 5616 AAAAAATTTTATGAATTTTGTATATAAATAATTTTTTTCTACTCT--GTTTGTAAACTATA 5674
Db 6316 TTTTACTTTATTTTTTATATATACAAATAAATTTTTGAAATTCATAAATAATGCAATGAAATA 6375
Qy 5675 TAAGAAAAAGATACATAACAGATCTTTTCAAGAGAAATCAATGGTTAGGAGAGAAATTTACT 5734
Db 6376 CATAAAAAATAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6435
Qy 5735 TTTTAAATCTGCATTTTAAAGAAATATAGATAGAGTTAGTTATTTTGAAGCAATCTTTA 5794
Db 6436 TATAATAATATATTTTTTCCGTGTTATTTATTTATCAITTTTTTTTTTGTATGCTATATATAT 6495
Qy 5795 TTTTATAGGAGAGGTATCTAAGTACAGTAAATTTCTTTTAAAGAGGTGTGTTTTTGGCA 5854
Db 6496 ATTATATAAATAATTAATAATAAACAACAAAAATTAATAAATAAATAAATAAATAAATAAATA 6555
Qy 5855 ATTGGAAAAATTTCAAAAAACAAGTATGTTTGTAAAGCAATAATATATGTCGAGGATTT 5914
Db 6556 AATATAATAACAATAAAGAAATATGATCTATATCAATTTATATATATATATATATATATATAT 6615
Qy 5915 TGACGTATCAATTTGTTAAAGATACTATACGTTGGCAAGTATTTTATTTATAGTTTACTAAT 5974
Db 6616 AATATATAGATAGATAAATATAGATAGAGAGAAACGAAGACATAATTTGCTCTCTTTGTTATC 6675
Qy 5975 GTTTAAATACGGAAAAACAGTCTAATTTTGTACAAATTTTAAATTTTGTAGAAATCTTTATAA 6034
Db 6676 TCTAATATATATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 6735
Qy 6035 AAAATATTTTAACTGTTTAAAGTATCTTAAACAAAAATTTCTTGTCTAATAAATTTTGTG 6094
Db 6736 AGTTTAATAATAATAATAAACAACGTTGCAATATATCTTTTTTATATGTTGTTATTTT 6795
Qy 6095 TATAAGAAATGTTTCGAAACAAAGTTTTTAAAAAATAATTTATGTTTATAATAGGAAGATAT 6154
Db 6796 CGTATTTTTTTTTTCTCAATTTATAATTTTACTTTAAATAAATAAATAAATAAATAAATAAATA 6855

| | | | |
|----|------|--------------------------------------------------------------|------|
| Qy | 6155 | CATGGATACT--ATTAGTAAATTTCTATAAATGTGACCTATATATAATGTAGAAAAATAT | 6212 |
| Db | 6856 | TATATATAATTAAATAGATAAAAGGATACATAAAATATAATTTCTGATTATATT | 6915 |
| Qy | 6213 | TTATCTAAATCTATAGTAGCATTTGTAATCAGACCTACAAACATATAGAGATTCTCTG | 6272 |
| Db | 6916 | TTTTTTTGTAGAAATTTTAAATTTATTATAAATTTTAAATATATATATATATATTTTTT | 6975 |
| Qy | 6273 | GTGAATGACGCTAGTAGCGATAATTCGGAAGAAATTTGTTTAGCATAT | 6320 |
| Db | 6976 | TTAAAAATATATAAACTAATAATTATTATTATATACATATTAAATAT | 7023 |

Search completed: April 28, 2005, 15:47:29
Job time : 1118 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2005, 14:17:37 ; Search time 3571 Seconds
(without alignments)
11919.659 Million cell updates/sec

Title: US-09-767-041-9
Perfect score: 6992
Sequence: 1 atcgccaaacgaattggca.....gcaaacagcaaaagagctt 6992

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5642217 seqs, 3043843248 residues

Total number of hits satisfying chosen parameters: 11284434

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:**

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 6992 | 100.0 | 6992 | 9 | US-09-767-041-9 |
| 2 | 6992 | 100.0 | 6992 | 9 | US-09-767-041-29 |
| 3 | 727.8 | 10.4 | 6865 | 9 | US-09-900-038A-3 |
| 4 | 727.8 | 10.4 | 6865 | 19 | US-10-865-873-3 |
| 5 | 726.8 | 10.4 | 17276 | 9 | US-09-870-759-83 |
| 6 | 726.8 | 10.4 | 17276 | 10 | US-09-751-708A-83 |
| 7 | 726.8 | 10.4 | 17276 | 18 | US-10-428-817A-79 |
| 8 | 721.4 | 10.3 | 25020 | 17 | US-10-192-280-1 |
| 9 | 684.8 | 9.8 | 18373 | 17 | US-10-461-990-4 |
| 10 | 652.4 | 9.3 | 2226 | 18 | US-10-804-408-166 |
| 11 | 652.4 | 9.3 | 2226 | 18 | US-10-804-408-167 |

| | | | | | | |
|------|-------|-----|---------|----|----------------------|-------------------|
| 12 | 650.8 | 9.3 | 2217 | 18 | US-10-804-408-162 | Sequence 162, App |
| 13 | 649.2 | 9.3 | 2226 | 18 | US-10-804-408-173 | Sequence 173, App |
| 14 | 648.8 | 9.3 | 2226 | 18 | US-10-804-408-171 | Sequence 171, App |
| 15 | 647.6 | 9.3 | 2217 | 18 | US-10-804-408-172 | Sequence 172, App |
| 16 | 647.6 | 9.3 | 2226 | 18 | US-10-804-408-168 | Sequence 168, App |
| 17 | 647.6 | 9.3 | 2226 | 18 | US-10-804-408-169 | Sequence 169, App |
| 18 | 646 | 9.2 | 2217 | 18 | US-10-804-408-164 | Sequence 164, App |
| 19 | 644.4 | 9.2 | 2217 | 18 | US-10-804-408-163 | Sequence 163, App |
| 20 | 644.4 | 9.2 | 2226 | 18 | US-10-804-408-170 | Sequence 170, App |
| 21 | 638.8 | 9.1 | 2225 | 18 | US-10-804-408-165 | Sequence 165, App |
| 22 | 428.4 | 6.1 | 6850 | 16 | US-10-182-960-1 | Sequence 1, Appli |
| 23 | 423.2 | 6.1 | 1368 | 9 | US-09-815-242-9587 | Sequence 9587, Ap |
| 24 | 419.8 | 6.0 | 2581 | 18 | US-10-804-408-178 | Sequence 178, App |
| 25 | 386 | 5.5 | 2577 | 18 | US-10-804-408-179 | Sequence 179, App |
| 26 | 336.6 | 4.8 | 2722 | 18 | US-10-804-408-176 | Sequence 176, App |
| 27 | 318.4 | 4.6 | 2692 | 18 | US-10-804-408-177 | Sequence 177, App |
| 28 | 301.4 | 4.3 | 8056 | 18 | US-10-473-126-386 | Sequence 386, App |
| 29 | 298.8 | 4.3 | 8056 | 18 | US-10-473-126-386 | Sequence 386, App |
| 30 | 222.2 | 3.2 | 8056 | 18 | US-10-473-126-240 | Sequence 240, App |
| C 31 | 219.8 | 3.1 | 8056 | 18 | US-10-473-126-240 | Sequence 240, App |
| C 32 | 206.6 | 3.0 | 50000 | 19 | US-10-706-635-23 | Sequence 23, Appl |
| C 33 | 205.2 | 2.9 | 3673778 | 16 | US-10-312-841-1 | Sequence 1, Appli |
| C 34 | 195.2 | 2.8 | 3673778 | 16 | US-10-312-841-1 | Sequence 1, Appli |
| 35 | 187.8 | 2.7 | 5979 | 14 | US-10-239-676-18 | Sequence 18, Appl |
| 36 | 187.8 | 2.7 | 5979 | 15 | US-10-240-453-26 | Sequence 26, Appl |
| 37 | 185.8 | 2.7 | 50000 | 19 | US-10-706-635-23 | Sequence 23, Appl |
| 38 | 183.6 | 2.6 | 8170 | 15 | US-10-240-453-132 | Sequence 132, App |
| 39 | 179 | 2.6 | 11691 | 15 | US-10-311-455-2214 | Sequence 2214, Ap |
| 40 | 176 | 2.5 | 372 | 17 | US-10-282-122A-38093 | Sequence 38093, A |
| 41 | 176 | 2.5 | 2337 | 18 | US-10-804-408-175 | Sequence 175, App |
| 42 | 174.6 | 2.5 | 3673778 | 16 | US-10-312-841-2 | Sequence 2, Appli |
| 43 | 172.8 | 2.5 | 15373 | 15 | US-10-311-455-440 | Sequence 440, App |
| 44 | 172.8 | 2.5 | 19634 | 18 | US-10-473-126-302 | Sequence 302, App |
| 45 | 171.4 | 2.5 | 2384 | 18 | US-10-804-408-174 | Sequence 174, App |

ALIGNMENTS

RESULT 1
US-09-767-041-9
; Sequence 9, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 6992
; TYPE: DNA
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(6992)
; OTHER INFORMATION: CPS 2
US-09-767-041-9

Query Match 100.0%; Score 6992; DB 9; Length 6992;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cyt 1 ATCCCAACGAATTGGCAATTATTGATATGATAGCAGTTGCAATTCTCAATCTTAA 60
|||||

Db 1 ATCGCCAAAGAAATTCGCATTTATTTGATGATAGCAGTTGCAATTTCTGCAATCTTAA 60
 Qy 61 CAAGTCATATACCAAAATCGTGATTTAAATCGTTCTGGAAATTTTATCATTAATGATGTTTC 120
 Db 61 CAAGTCATATACCAAAATCGTGATTTAAATCGTTCTGGAAATTTTATCATTAATGATGTTTC 120
 Qy 121 ATTATTTTGGCAATTTTATATCTCGTATGCCAGTTGGAATTTGAGTATAGAGGTAATCTGA 180
 Db 121 ATTATTTTGGCAATTTTATATCTCGTATGCCAGTTGGAATTTGAGTATAGAGGTAATCTGA 180
 Qy 181 TAGAGTTTGGAAAAACATTTAACTATAGTATAATTTTGCAATTTTCTTACGGCAGTAT 240
 Db 181 TAGAGTTTGGAAAAACATTTAACTATAGTATAATTTTGCAATTTTCTTACGGCAGTAT 240
 Qy 241 CATTTTGTGGAGAAATAATTTCCGACCTTTCAGACGCTGGTGGCGGTATTTTACATPAA 300
 Db 241 CATTTTGTGGAGAAATAATTTCCGACCTTTCAGACGCTGGTGGCGGTATTTTACATPAA 300
 Qy 301 TAAACTTCGTTTGGTATACCTATTTAACTGTAATTTTAAAGCAGTTTAAAGTAGCTTTC 360
 Db 301 TAAACTTCGTTTGGTATACCTATTTAACTGTAATTTTAAAGCAGTTTAAAGTAGCTTTC 360
 Qy 361 TATTTTCCGACAAATCTATCAAAAAAGACGATTCTAAATTAACAACGCTGAACGATGGGAAA 420
 Db 361 TATTTTCCGACAAATCTATCAAAAAAGACGATTCTAAATTAACAACGCTGAACGATGGGAAA 420
 Qy 421 ATATGCAAGTTTATTTGGAATCAATAAACAAATTTCAAAAAATCTTTGTTGCAATGGTAG 480
 Db 421 ATATGCAAGTTTATTTGGAATCAATAAACAAATTTCAAAAAATCTTTGTTGCAATGGTAG 480
 Qy 481 TTTTAGGTACAGAAATAGATAAATTAATTTATCAATTTACCGCTCTATTTCTGTGGAAG 540
 Db 481 TTTTAGGTACAGAAATAGATAAATTAATTTATCAATTTACCGCTCTATTTCTGTGGAAG 540
 Qy 541 AAGCTATAGAGTTTTCACAAAGGAAGTGGTCGACACAGCTCTTTATAAATCTACCAAGTG 600
 Db 541 AAGCTATAGAGTTTTCACAAAGGAAGTGGTCGACACAGCTCTTTATAAATCTACCAAGTG 600
 Qy 601 AGTTTGTAGAGTTTAAAGCAATTCGTTTTCAGATTTTGAAGTTTGTAGGTAATGATGAAGCG 660
 Db 601 AGTTTGTAGAGTTTAAAGCAATTCGTTTTCAGATTTTGAAGTTTGTAGGTAATGATGAAGCG 660
 Qy 661 TTGATATTAATTCATTCGTTTACTCGGTTTGAAGAAACAAAAAATCCAACTGCTAGGTTG 720
 Db 661 TTGATATTAATTCATTCGTTTACTCGGTTTGAAGAAACAAAAAATCCAACTGCTAGGTTG 720
 Qy 721 ACCATAGCATTTGTAACCTTTTCCAAATTTTATAAGCCTAGTCATATCATGATGAAC 780
 Db 721 ACCATAGCATTTGTAACCTTTTCCAAATTTTATAAGCCTAGTCATATCATGATGAAC 780
 Qy 781 GACTTTTGGATATACTCGGAGCGGTAGTCGGGTTTAAATTTTGTGGTATAGTTTCTATTT 840
 Db 781 GACTTTTGGATATACTCGGAGCGGTAGTCGGGTTTAAATTTTGTGGTATAGTTTCTATTT 840
 Qy 841 TGTAGTTTCCAAATTTTGTAGATGGTCGACCGGCTATTTTGTCTCAGAAACGAGTTG 900
 Db 841 TGTAGTTTCCAAATTTTGTAGATGGTCGACCGGCTATTTTGTCTCAGAAACGAGTTG 900
 Qy 901 GACAGATGGACGATATTTTACATTTCTACAGTTTTCGATCGATGATGTTGATGCTCAGG 960
 Db 901 GACAGATGGACGATATTTTACATTTCTACAGTTTTCGATCGATGATGTTGATGCTCAGG 960
 Qy 961 AGCGCAAAAAAGACTTGTACGCCAAACCCAGATGCAAGGGTGGGTATGTTTTAAATGG 1020
 Db 961 AGCGCAAAAAAGACTTGTACGCCAAACCCAGATGCAAGGGTGGGTATGTTTTAAATGG 1020
 Qy 1021 GAAAAACGATCCTAGAAATTTACTCCAAATTTGACATTTTATAGCAGGTTTGAAGT 1080
 Db 1021 GAAAAACGATCCTAGAAATTTACTCCAAATTTGACATTTTATAGCAGGTTTGAAGT 1080
 Qy 1081 AGTTTACCACAGTTTATTAATGTTTAAATGTCGATATGAGTCTAGTTGTTGACGCTCCAC 1140
 Db 1081 AGTTTACCACAGTTTATTAATGTTTAAATGTCGATATGAGTCTAGTTGTTGACGCTCCAC 1140

Qy 1141 CTACAGTTGATGAAATTTGAAAAATATATCTCTCGTCAAAAGAGACGATTCGAGTTTAAAC 1200
 Db 1141 CTACAGTTGATGAAATTTGAAAAATATATCTCTCGTCAAAAGAGACGATTCGAGTTTAAAC 1200
 Qy 1201 CAGGGATTAACAGTCTCTGCGAGGTTAGTGGTCTGATGATATATCAGACGCTTCGACGACG 1260
 Db 1201 CAGGGATTAACAGTCTCTGCGAGGTTAGTGGTCTGATGATATATCAGACGCTTCGACGACG 1260
 Qy 1261 TAGTTCGGTTGGACTTTAGCATACATTTGATAAATTTGACTATCTGGTCTGATATTTAAATTT 1320
 Db 1261 TAGTTCGGTTGGACTTTAGCATACATTTGATAAATTTGACTATCTGGTCTGATATTTAAATTT 1320
 Qy 1321 TATTAAGACAGTGAAGTTGTTTGTGAGAGAGGGAAGTAAAGTATATGAAAG 1380
 Db 1321 TATTAAGACAGTGAAGTTGTTTGTGAGAGAGGGAAGTAAAGTATATGAAAG 1380
 Qy 1381 TTTGTTTGGTTCGGTCTCTCAGGGGACATTTGACTCACTGTATTTGTTAAAAACCGTTT 1440
 Db 1381 TTTGTTTGGTTCGGTCTCTCAGGGGACATTTGACTCACTGTATTTGTTAAAAACCGTTT 1440
 Qy 1441 GGAAGGAAGAAACGTTTTTGGTTAAACATTTGATAAAGAGGATGCAAGAGTCTTTTGA 1500
 Db 1441 GGAAGGAAGAAACGTTTTTGGTTAAACATTTGATAAAGAGGATGCAAGAGTCTTTTGA 1500
 Qy 1501 AGAATGAAAAAATGTATCCATGTTACTTTCCAAACAAATCGCAATCTCAATTAATTTAGTGA 1560
 Db 1501 AGAATGAAAAAATGTATCCATGTTACTTTCCAAACAAATCGCAATCTCAATTAATTTAGTGA 1560
 Qy 1561 AAAATCTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACAGATGTTATTTTTCAT 1620
 Db 1561 AAAATCTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACAGATGTTATTTTTCAT 1620
 Qy 1621 CTGGTCGGCCGTTGCTGCTCCCTTTTACATCGGAAAACTATTTTGGAGCAAGACGA 1680
 Db 1621 CTGGTCGGCCGTTGCTGCTCCCTTTTACATCGGAAAACTATTTTGGAGCAAGACGA 1680
 Qy 1681 TTTATATTTGAAGTATTTGATTCGAGTTTAAATAAATCTACATTTAACTGGAATACTAGTTTATC 1740
 Db 1681 TTTATATTTGAAGTATTTGATTCGAGTTTAAATAAATCTACATTTAACTGGAATACTAGTTTATC 1740
 Qy 1741 CCGTAACAGATATTTTATTTGTTTCACTGGGAAGAAATGAAGGTATATCTTAAATCTA 1800
 Db 1741 CCGTAACAGATATTTTATTTGTTTCACTGGGAAGAAATGAAGGTATATCTTAAATCTA 1800
 Qy 1801 TTAACCTTGGGAGTATTTTATTTTAAATGATTTTGTAAACAGTAGGAACTCATGAAACAGTT 1860
 Db 1801 TTAACCTTGGGAGTATTTTATTTTAAATGATTTTGTAAACAGTAGGAACTCATGAAACAGTT 1860
 Qy 1861 TAATCGATTTGATAAAAGAGATTCATTTATTTGAAAAAATTTGGAAGTATAACCGACGAAAT 1920
 Db 1861 TAATCGATTTGATAAAAGAGATTCATTTATTTGAAAAAATTTGGAAGTATAACCGACGAAAT 1920
 Qy 1921 ATTATTTCAACAGGATTTCTGACTATATTTCCAGAAATTTGCAAGTATATAAATAATTTCT 1980
 Db 1921 ATTATTTCAACAGGATTTCTGACTATATTTCCAGAAATTTGCAAGTATATAAATAATTTCT 1980
 Qy 1981 CAGTTACAAAGAAATGGAACAAATATTAACAATCAGAAATGAGTTATTTTCCACCGGAGG 2040
 Db 1981 CAGTTACAAAGAAATGGAACAAATATTAACAATCAGAAATGAGTTATTTTCCACCGGAGG 2040
 Qy 2041 CCCCGTACTTTTATGAAATTCATTTATCCAAAGGAAAAAACAATTTATGTTCTAGACA 2100
 Db 2041 CCCCGTACTTTTATGAAATTCATTTATCCAAAGGAAAAAACAATTTATGTTCTAGACA 2100
 Qy 2101 AAAAAAGTATGGTGAACATGTAATCATCATCAGTAGAGTTTGTAAAGAAATTTTACA 2160
 Db 2101 AAAAAAGTATGGTGAACATGTAATCATCATCAGTAGAGTTTGTAAAGAAATTTTACA 2160
 Qy 2161 AGATAATAATATTTTATTTATGAAAAATATAGATGATTTGTTGAAAAAATTTATGAGT 2220
 Db 2161 AGATAATAATATTTTATTTATGAAAAATATAGATGATTTGTTGAAAAAATTTATGAGT 2220

| | | | |
|----|------|----------------------------------------------------------------|------|
| QY | 2221 | TTCTAGCAAACTAACCTTTACATCAAAATAATAATTTTTTTTGTGAAAGATTAAACAAAT | 2280 |
| DB | 2221 | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT | 2280 |
| QY | 2281 | AGTTGAAAAATTTAAATGAGATCAAGAAATGAATAATAAAAAAGATGCATATTTTCATAA | 2340 |
| DB | 2281 | AGTTGAAAAATTTAAATGAGATCAAGAAATGAATAATAAAAAAGATGCATATTTTCATAA | 2340 |
| QY | 2341 | TGGCTTATCATAAATTTTTCTCAGATTTTACTGGAGAGGATACAGATATTTATCATCTTCT | 2400 |
| DB | 2341 | TGGCTTATCATAAATTTTTCTCAGATTTTACTGGAGAGGATACAGATATTTATCATCTTCT | 2400 |
| QY | 2401 | CTCAGAGATGCACACCATTTAGTCTTCAGATACCTGTATAATTTTAAATATTC | 2460 |
| DB | 2401 | CTCAGAGATGCACACCATTTAGTCTTCAGATACCTGTATAATTTTAAATATTC | 2460 |
| QY | 2461 | TCAGGATTTATATGTTGAAATTTACAAAAGATGACAAAATAATAAGAAAAATAGGATATA | 2520 |
| DB | 2461 | TCAGGATTTATATGTTGAAATTTACAAAAGATGACAAAATAATAAGAAAAATAGGATATA | 2520 |
| QY | 2521 | TGAACGAGTTAAATGTTACAGATTTATTTCTTAATATATCAGAAAAAATTTGATTAATGT | 2580 |
| DB | 2521 | TGAACGAGTTAAATGTTACAGATTTATTTCTTAATATATCAGAAAAAATTTGATTAATGT | 2580 |
| QY | 2581 | ACTGTTTAGAATTTTAAATGAGATGATCGAGCTTTTGAATACATTTTACAAGATTTGT | 2640 |
| DB | 2581 | ACTGTTTAGAATTTTAAATGAGATGATCGAGCTTTTGAATACATTTTACAAGATTTGT | 2640 |
| QY | 2641 | GTTTATGATAGATTAAGAAATGCTCTAAGATTAAGATTTGTTCTAATTTGGGTTTCG | 2700 |
| DB | 2641 | GTTTATGATAGATTAAGAAATGCTCTAAGATTAAGATTTGTTCTAATTTGGGTTTCG | 2700 |
| QY | 2701 | CTTCCACATGATTTTGTGGCAATCTTTTATCAAAATGAAAAACGAAACAGCTTATTTATTT | 2760 |
| DB | 2701 | CTTCCACATGATTTTGTGGCAATCTTTTATCAAAATGAAAAACGAAACAGCTTATTTATTT | 2760 |
| QY | 2761 | AAGTAATCTAAATGTCAGATGAATCTATTATACAGCAAAATATAGAAAAATATGAATTT | 2820 |
| DB | 2761 | AAGTAATCTAAATGTCAGATGAATCTATTATACAGCAAAATATAGAAAAATATGAATTT | 2820 |
| QY | 2821 | TCAAATAGATTTCTAAATGGAATTTAAGATATATAAGTGGAAAAAATCAACATCT | 2880 |
| DB | 2821 | TCAAATAGATTTCTAAATGGAATTTAAGATATATAAGTGGAAAAAATCAACATCT | 2880 |
| QY | 2881 | TCCTCTATTGTCCTTACAGATGATCTATTGATGAATGCTAAATGCAAGAAATTTAGGT | 2940 |
| DB | 2881 | TCCTCTATTGTCCTTACAGATGATCTATTGATGAATGCTAAATGCAAGAAATTTAGGT | 2940 |
| QY | 2941 | TTTTTATTTGCTAGAAAGTTTAAAAATAGAAAAATAAATCTAAATTTTAAAGAAATTTTACT | 3000 |
| DB | 2941 | TTTTTATTTGCTAGAAAGTTTAAAAATAGAAAAATAAATCTAAATTTTAAAGAAATTTTACT | 3000 |
| QY | 3001 | AAAAATATAATAGTTGATTTGAGAGTAATGTATGTTTAAATTTTAAATATATGACC | 3060 |
| DB | 3001 | AAAAATATAATAGTTGATTTGAGAGTAATGTATGTTTAAATTTTAAATATATGACC | 3060 |
| QY | 3061 | CGGAATATTTTATTTTAACTCTCTGTTGATTTTATTTTATTCAGAGCAAAAGATG | 3120 |
| DB | 3061 | CGGAATATTTTATTTTAACTCTCTGTTGATTTTATTTTATTCAGAGCAAAAGATG | 3120 |
| QY | 3121 | TATTTTATTTAAATTTTATGAAATTTTAAATTTTATTTTATTTTAAACTAAGC | 3180 |
| DB | 3121 | TATTTTATTTAAATTTTATGAAATTTTAAATTTTATTTTATTTTAAACTAAGC | 3180 |
| QY | 3181 | TAAATATAAAAAATGAAATTTTATGTTTATTAATGCTATATATGTTTGTTCAG | 3240 |
| DB | 3181 | TAAATATAAAAAATGAAATTTTATGTTTATTAATGCTATATATGTTTGTTCAG | 3240 |
| QY | 3241 | TAGTCACAGTATGTTTGTGAAATTAATTTTGAAGATTTTTCAGATTTTACTGCTC | 3300 |
| DB | 3241 | TAGTCACAGTATGTTTGTGAAATTAATTTTGAAGATTTTTCAGATTTTACTGCTC | 3300 |
| QY | 3301 | CCATAATTTGGAATTTTGAATATGATTTATTAATTTGTTTCTTATTTATTAATTTGATT | 3360 |

| | | | |
|----|------|-------------------------------------------------------------------|------|
| DB | 3301 | CCATAATTTGGAATTTTGAATATGATTTATTAATTTGTTTCTTATTTATAATTTGATT | 3360 |
| QY | 3361 | ATAAAAAATAAAAAATAGTATCTTTTTTAGTTTTTTTAGTTTTTTTAGGTATATCTGCAT | 3420 |
| DB | 3361 | ATAAAAAATAAAAAATAGTATCTTTTTTAGTTTTTTTAGTTTTTTTAGGTATATCTGCAT | 3420 |
| QY | 3421 | TGTATATATTTCAAAATGGGAAAGATTTGTATTTTTTAGACACAGACCTTATAGGACTAG | 3480 |
| DB | 3421 | TGTATATATTTCAAAATGGGAAAGATTTGTATTTTTTAGACACAGACCTTATAGGACTAG | 3480 |
| QY | 3481 | ACTATCTTATAACAGCGCTCAAAAACAAAGTTGCTGCTTTTATGAACACTATCTACGTAA | 3540 |
| DB | 3481 | ACTATCTTATAACAGCGCTCAAAAACAAAGTTGCTGCTTTTATGAACACTATCTACGTAA | 3540 |
| QY | 3541 | ATACCACTACAATTTATAGTTTCCAAATTCCTTAAATCTTTGCACCTTATAAAAAATAAAATGC | 3600 |
| DB | 3541 | ATACCACTACAATTTATAGTTTCCAAATTCCTTAAATCTTTGCACCTTATAAAAAATAAAATGC | 3600 |
| QY | 3601 | AACAAATTTTTTCTGCTCTGCTTTTATACCGATCTTATTAAGTGGATCGAGAAATTCG | 3660 |
| DB | 3601 | AACAAATTTTTTCTGCTCTGCTTTTATACCGATCTTATTAAGTGGATCGAGAAATTCG | 3660 |
| QY | 3661 | GTAGTTTATCGCTAGCAATTAATTAATGCTTGTATGAGATATATAGGTGGAAAAAT | 3720 |
| DB | 3661 | GTAGTTTATCGCTAGCAATTAATTAATGCTTGTATGAGATATATAGGTGGAAAAAT | 3720 |
| QY | 3721 | TTGCTTGGATAAAAAGCTTAATAGTAAATTTTGTAAATCTACTTATTTTAAATACTG | 3780 |
| DB | 3721 | TTGCTTGGATAAAAAGCTTAATAGTAAATTTTGTAAATCTACTTATTTTAAATACTG | 3780 |
| QY | 3781 | AAATGCTTTACCATGAAATTTTGGCTGTTTATAATTTCTAGAGAAATCAAGTAACGAAGCTA | 3840 |
| DB | 3781 | AAATGCTTTACCATGAAATTTTGGCTGTTTATAATTTCTAGAGAAATCAAGTAACGAAGCTA | 3840 |
| QY | 3841 | GAATTTATTTTATCAAGGAATTTTGAATAAGTATTAGAAAACAATATTTTATTTGGAT | 3900 |
| DB | 3841 | GAATTTATTTTATCAAGGAATTTTGAATAAGTATTAGAAAACAATATTTTATTTGGAT | 3900 |
| QY | 3901 | ATGGAATATCCGAATATTCAGTTTACGGGAATTTGGCTCGGAAGCTATTTCAGGCTATATAT | 3960 |
| DB | 3901 | ATGGAATATCCGAATATTCAGTTTACGGGAATTTGGCTCGGAAGCTATTTCAGGCTATATAT | 3960 |
| QY | 3961 | CAATTTTTTATTAATCAGGAATAGTTGGGTGATTTTACTGATGTTTTCTTTTTTATG | 4020 |
| DB | 3961 | CAATTTTTTATTAATCAGGAATAGTTGGGTGATTTTACTGATGTTTTCTTTTTTATG | 4020 |
| QY | 4021 | TTATAAAAAAAGTTATGAGTTTAAATGGGAAAAACAGACTATTTTATTTACATCATTAG | 4080 |
| DB | 4021 | TTATAAAAAAAGTTATGAGTTTAAATGGGAAAAACAGACTATTTTATTTACATCATTAG | 4080 |
| QY | 4081 | CCATATTTTTTCATATATGAAACAAATAGATCCGATTTATTTATTTAGTACTATTTCTTTT | 4140 |
| DB | 4081 | CCATATTTTTTCATATATGAAACAAATAGATCCGATTTATTTATTTAGTACTATTTCTTTT | 4140 |
| QY | 4141 | CTTCAATAGGTATTTTGGAAATAATATAATTTTAAAAAGGATATGGAGACAAAAATGAAT | 4200 |
| DB | 4141 | CTTCAATAGGTATTTTGGAAATAATATAATTTTAAAAAGGATATGGAGACAAAAATGAAT | 4200 |
| QY | 4201 | GAATTAATTTTCAGTTTATGTA CCAATTTTATAATGTC CCAAGATTTATCTTGATAATGTAAT | 4260 |
| DB | 4201 | GAATTAATTTTCAGTTTATGTA CCAATTTTATAATGTC CCAAGATTTATCTTGATAATGTAAT | 4260 |
| QY | 4261 | AACAGTATTTTAA CCAACATATACTAATTTAGAGTTTATCTCCGTAATGATGGAAT | 4320 |
| DB | 4261 | AACAGTATTTTAA CCAACATATACTAATTTAGAGTTTATCTCCGTAATGATGGAAT | 4320 |
| QY | 4321 | ACTGATGATTTCTGAGAAAAATTTGCTTAAATATATGAAACGATGGAAGAAATTAATAT | 4380 |
| DB | 4321 | ACTGATGATTTCTGAGAAAAATTTGCTTAAATATATGAAACGATGGAAGAAATTAATAT | 4380 |
| QY | 4381 | TACAAGAAAAATTAATCGCGGTCTAGCAGATGCTCGAAATTTTCGGACTAGAACATGCAACA | 4440 |

4381 TAACAAGAAATTAATGGCGGTCTAGCAGATGCTCGAAATTTCCGGACTAGAAACATGCAACA 4440 Db
4441 GGTAAATATATGCTTTTGTGCGATTCCTGATGACTATATAGAAGTTGCCAATGTTCCAGAGA 4500 Qy
4441 GGTAAATATATGCTTTTGTGCGATTCCTGATGACTATATAGAAGTTGCCAATGTTCCAGAGA 4500 Db
4501 ATGCATGATAATTAACCTCAGTATATAATCGCGATATAGCAGATAGATTTTTTGTAGTA 4560 Qy
4501 ATGCATGATAATTAACCTCAGTATATAATCGCGATATAGCAGATAGATTTTTTGTAGTA 4560 Db
4561 GACGAAACCGGTPATCAAGAAAGAAAGAAATAGTAAATTTTCATGCTTTAACGAGAGAA 4620 Qy
4561 GACGAAACCGGTPATCAAGAAAGAAAGAAATAGTAAATTTTCATGCTTTAACGAGAGAA 4620 Db
4621 GAGACTGTAAGAGAAATTTTGTGAGATCTAATATAGAAATTAAGTTTGGTGGCAGCTT 4680 Qy
4621 GAGACTGTAAGAGAAATTTTGTGAGATCTAATATAGAAATTAAGTTTGGTGGCAGCTT 4680 Db
4681 TATTACAGAGATATTAAGAGATATAAATTTCCAAATTAATAAGAGATTTGGTGAG 4740 Qy
4681 TATTACAGAGATATTAAGAGATATAAATTTCCAAATTAATAAGAGATTTGGTGAG 4740 Db
4741 GATTTGCTTTTAAATTTGGAGTCTTTGAACAATGTAAACACGTTAGTCTGATCTAGA 4800 Qy
4741 GATTTGCTTTTAAATTTGGAGTCTTTGAACAATGTAAACACGTTAGTCTGATCTAGA 4800 Db
4801 GAATATATTAATTAATGATGCTGTAACAGTTCGCTTATTAATCAGAAATTCCTATA 4860 Qy
4801 GAATATATTAATTAATGATGCTGTAACAGTTCGCTTATTAATCAGAAATTCCTATA 4860 Db
4861 AATAATATTGATTTAGTCAACAGATTTGAGAAATACCCCTTTAAAGTTAAAGAGAGATTT 4920 Qy
4861 AATAATATTGATTTAGTCAACAGATTTGAGAAATACCCCTTTAAAGTTAAAGAGAGATTT 4920 Db
4921 AGTCATTATTTGATCAAAAGTTATTAAGAGAGAGTTAAATGTTTAAACAAAATGTAT 4980 Qy
4921 AGTCATTATTTGATCAAAAGTTATTAAGAGAGAGTTAAATGTTTAAACAAAATGTAT 4980 Db
4981 TCACAGATGTTTGGATTAATGAGTTCTTGGCCAAATATTAAGTCTTTATCGAAAGAAATA 5040 Qy
4981 TCACAGATGTTTGGATTAATGAGTTCTTGGCCAAATATTAAGTCTTTATCGAAAGAAATA 5040 Db
5041 CGTAGATATCCATTTAATAAGCGAAGATATTTATCAAGAACATTTTACGTTG 5100 Qy
5041 CGTAGATATCCATTTAATAAGCGAAGATATTTATCAAGAACATTTTACGTTG 5100 Db
5101 TATTTGATGAAATTTTCGCTAAACTATATGTAATGTTATATAAGAAATTTCAAAGCAG 5160 Qy
5101 TATTTGATGAAATTTTCGCTAAACTATATGTAATGTTATATAAGAAATTTCAAAGCAG 5160 Db
5161 TAGAGTAAATGGATAAATTTAGTGTATTTGTTCCAGTTTATATGTAGATAAATATT 5220 Qy
5161 TAGAGTAAATGGATAAATTTAGTGTATTTGTTCCAGTTTATATGTAGATAAATATT 5220 Db
5221 TAAGTAGTTGTATAGAAACATTTATTAATCAAAATTTATAAATATAGAAATATTATGA 5280 Qy
5221 TAAGTAGTTGTATAGAAACATTTATTAATCAAAATTTATAAATATAGAAATATTATGA 5280 Db
5281 TAGATGATGGCTCTGTAGATGATTTCTGCTAAATATGCAAGGAATATGAGAAAGATA 5340 Qy
5281 TAGATGATGGCTCTGTAGATGATTTCTGCTAAATATGCAAGGAATATGAGAAAGATA 5340 Db
5341 AAAGAGTAAATTTTTTTCATTAATCATAGTGGAGTATCAATGCTGAAATCATGGAA 5400 Qy
5341 AAAGAGTAAATTTTTTTCATTAATCATAGTGGAGTATCAATGCTGAAATCATGGAA 5400 Db
5401 TAAAGCGAGTACAGCTGAATATATATGTTGTTGACCTGATGATGTTGTTGATGTA 5460 Qy
5401 TAAAGCGAGTACAGCTGAATATATATGTTGTTGACCTGATGATGTTGTTGATGTA 5460 Db
5461 GATTAGTAGAAATTTATATTTTAAATTTATAAAAAGTAGAAGTATCTGTTGTT 5520 Qy
5461 GATTAGTAGAAATTTATATTTTAAATTTATAAAAAGTAGAAGTATCTGTTGTT 5520 Db
5461 GATTAGTAGAAATTTATATTTTAAATTTATAAAAAGTAGAAGTATCTGTTGTT 5520 Db

5521 TGTAGCTACTTTTTCAGAAATATATAATTAATTTTGAAGTGAATATCCAAATATTGATT 5580 Qy
5521 TGTAGCTACTTTTTCAGAAATATATAATTAATTTTGAAGTGAATATCCAAATATTGATT 5580 Db
5581 TTGAAGCAATTAATACCGTGCAGGACATGGGAGAAAAATTTTATGAATTTGTATATAA 5640 Qy
5581 TTGAAGCAATTAATACCGTGCAGGACATGGGAGAAAAATTTTATGAATTTGTATATAA 5640 Db
5641 ATAATATTTTCTACTCTCTCTGTTTAAACTATATAAGAAAAAGATACATAACAGATCTTT 5700 Qy
5641 ATAATATTTTCTACTCTCTCTGTTTAAACTATATAAGAAAAAGATACATAACAGATCTTT 5700 Db
5701 TTCAAGAGATCAATCGTTAGGAGAGATTTACTTTTTTAACTGCAATTTTAAAGAATA 5760 Qy
5701 TTCAAGAGATCAATCGTTAGGAGAGATTTACTTTTTTAACTGCAATTTTAAAGAATA 5760 Db
5761 TAGATAGAGTTAGTTTATTTGACTGAAATCTTTTATATAGGAGAGGTATCTAAGTA 5820 Qy
5761 TAGATAGAGTTAGTTTATTTGACTGAAATCTTTTATATAGGAGAGGTATCTAAGTA 5820 Db
5821 CAGTAAATTTCTTTTAAAGAGAGGTGTGTTTTTGCATTCGAAAAATTTGCAAAAAACAAGTGA 5880 Qy
5821 CAGTAAATTTCTTTTAAAGAGAGGTGTGTTTTTGCATTCGAAAAATTTGCAAAAAACAAGTGA 5880 Db
5881 TAGTATGTTTAAAGCAATATATGTTGAGGATTTTGACGTATCAATTTGTTAAAGATACTA 5940 Qy
5881 TAGTATGTTTAAAGCAATATATGTTGAGGATTTTGACGTATCAATTTGTTAAAGATACTA 5940 Db
5941 TACGTTGGCAAGTATTTTATTTATAGCTTACTAACTGTTTAAATACGGAACAGTCTATT 6000 Qy
5941 TACGTTGGCAAGTATTTTATTTATAGCTTACTAACTGTTTAAATACGGAACAGTCTATT 6000 Db
6001 TTGACAAATTTTAAATTTTAAAGAAATCTTTTAAAGAAATATTTTAACTTTGTTAAAG 6060 Qy
6001 TTGACAAATTTTAAATTTTAAAGAAATCTTTTAAAGAAATATTTTAACTTTGTTAAAG 6060 Db
6061 TATCTAACAAAAATTTCTTTGCTTAAAAAAATTTTGTATAGAAATTTGTTGCAACAAGTTT 6120 Qy
6061 TATCTAACAAAAATTTCTTTGCTTAAAAAAATTTTGTATAGAAATTTGTTGCAACAAGTTT 6120 Db
6121 TTAAGAAAAATATATGTTTATAAGAGATATCATCGATACTATAGTAAAAATTTCTA 6180 Qy
6121 TTAAGAAAAATATATGTTTATAAGAGATATCATCGATACTATAGTAAAAATTTCTA 6180 Db
6181 TAAATGTACCTATATATAATGTAGAAAAATTTTATCTAAATGTATAGATGCAATTTGTA 6240 Qy
6181 TAAATGTACCTATATATAATGTAGAAAAATTTTATCTAAATGTATAGATGCAATTTGTA 6240 Db
6241 ATCAGACCTACAAACATATAGAGATTTCTTCTGGTGAATGACGGTAGTACGGATAATTCGG 6300 Qy
6241 ATCAGACCTACAAACATATAGAGATTTCTTCTGGTGAATGACGGTAGTACGGATAATTCGG 6300 Db
6301 AAGAAATTTGTTTACATATCGAAGAAAGATAGTCGCAATTCGTTATTTTAAAAAGAGA 6360 Qy
6301 AAGAAATTTGTTTACATATCGAAGAAAGATAGTCGCAATTCGTTATTTTAAAAAGAGA 6360 Db
6361 ACGCGGGCTATCAGATGCCCGTAAATTTATGTCATAGTTCGCGCCCAAGGCTACTCTTAG 6420 Qy
6361 ACGCGGGCTATCAGATGCCCGTAAATTTATGTCATAGTTCGCGCCCAAGGCTACTCTTAG 6420 Db
6421 CTTTATAGACTCAGATGATTTTATTTCAATTCGGAGTTTATCCAACTGTTTACACGAGCA 6480 Qy
6421 CTTTATAGACTCAGATGATTTTATTTCAATTCGGAGTTTATCCAACTGTTTACACGAGCA 6480 Db
6481 TTGAGAGAGAAATGCCCTTGTGGCAGTTGCTGTTATAGTAGGGTAGATGTTTCGGGCG 6540 Qy
6481 TTGAGAGAGAAATGCCCTTGTGGCAGTTGCTGTTATAGTAGGGTAGATGTTTCGGGCG 6540 Db
6541 ATTTCTTAACAGCAGCAGCGCTTCTTCAAAATCAGGCTGTTCTGAGCGGCGGAATGTTT 6600 Qy
6541 ATTTCTTAACAGCAGCAGCGCTTCTTCAAAATCAGGCTGTTCTGAGCGGCGGAATGTTT 6600 Db

Db 1261 TAGTTGGTGGACTTAGCATACATTCGATTAATTCGACTATCTCGTCAGATATTAAATTT 1320
Qy 1321 TATTAAAGACAGTGAAGTTGTATTTGTGTGAGAGAGGGAAGTAAGTAAAGTATATGAAAG 1380
Db 1321 TATTAAAGACAGTGAAGTTGTATTTGTGTGAGAGAGGGAAGTAAGTAAAGTATATGAAAG 1380
Qy 1381 TTTGTGTTGGTCCGTTCTTTCAGGGGACATTTGACTCACTCTGTTATTTCTTAAACCGTTTT 1440
Db 1381 TTTGTGTTGGTCCGTTCTTTCAGGGGACATTTGACTCACTCTGTTATTTCTTAAACCGTTTT 1440
Qy 1441 GGAAGGAAGAAACGTTTTTGGGTAAACATTTGATAAAGAGGATGCAAGAACTCTTTTGA 1500
Db 1441 GGAAGGAAGAAACGTTTTTGGGTAAACATTTGATAAAGAGGATGCAAGAACTCTTTTGA 1500
Qy 1501 AGAATGAAAAATGTATCCATGTTACTTTCCAAACAAATCGCAATCTCATTTAATTTAGTGA 1560
Db 1501 AGAATGAAAAATGTATCCATGTTACTTTCCAAACAAATCGCAATCTCATTTAATTTAGTGA 1560
Qy 1561 AAAATACCTTCTTAGCTTCAAAAATTTTACGTGATGAGAACCAAGATGTTATTTATTCAT 1620
Db 1561 AAAATACCTTCTTAGCTTCAAAAATTTTACGTGATGAGAACCAAGATGTTATTTATTCAT 1620
Qy 1621 CTGCTGGCGCGTTCGTCTCCCTTTCTTTTACATCGGAAACTATTTGGAGCAAGACGA 1680
Db 1621 CTGCTGGCGCGTTCGTCTCCCTTTCTTTTACATCGGAAACTATTTGGAGCAAGACGA 1680
Qy 1681 TTTATATGAAAGTATTTGATCGAGTTAATAAATCTACATTAACCTGGAATACTAGTTTATC 1740
Db 1681 TTTATATGAAAGTATTTGATCGAGTTAATAAATCTACATTAACCTGGAATACTAGTTTATC 1740
Qy 1741 CCGTAACAGATATTTTATTTGTTTCAGTGGGAGGAATGAGNAGGTATATCTTAATCTA 1800
Db 1741 CCGTAACAGATATTTTATTTGTTTCAGTGGGAGGAATGAGNAGGTATATCTTAATCTA 1800
Qy 1801 TTAATCTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAATCATGAACACAGTT 1860
Db 1801 TTAATCTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAATCATGAACACAGTT 1860
Qy 1861 TAATCGATTGATAAAGAGATTGATTTTATTTGAAAAAAAATGGAAGTATAACCGACGAAT 1920
Db 1861 TAATCGATTGATAAAGAGATTGATTTTATTTGAAAAAAAATGGAAGTATAACCGACGAAT 1920
Qy 1921 ATTTTATCAACAGGATATCTGACTATATTCAGATATTTGCAAGTATAAAAAATTTCT 1980
Db 1921 ATTTTATCAACAGGATATCTGACTATATTCAGATATTTGCAAGTATAAAAAATTTCT 1980
Qy 1981 CAGTTACAAAGAAATGGAACAAATATATTAACAAATCAGAACTAGTTATTTGCCACGGAG 2040
Db 1981 CAGTTACAAAGAAATGGAACAAATATATTAACAAATCAGAACTAGTTATTTGCCACGGAG 2040
Qy 2041 CCCCCTACTTTTATGAATTCATTTATCAAGGAAAAAAAACAAATATTTGTTCTAGACA 2100
Db 2041 CCCCCTACTTTTATGAATTCATTTATCAAGGAAAAAAAACAAATATTTGTTCTAGACA 2100
Qy 2101 AAAAAAGTATGCTGAACATGTAATGATCATCAAGTAGAGTTTGTAGAGAAATTTTACA 2160
Db 2101 AAAAAAGTATGCTGAACATGTAATGATCATCAAGTAGAGTTTGTAGAGAAATTTTACA 2160
Qy 2161 AGATAAATAATTTTATTTATAGAAAAATATAGATGATTTGTTGAAAAAATTTATGAAAT 2220
Db 2161 AGATAAATAATTTTATTTATAGAAAAATATAGATGATTTGTTGAAAAAATTTATGAAAT 2220
Qy 2221 TTCTAAGCAAACTAACTTACATCAAAATAATTTTTTTTGTGAAAGATTTAAACAAAT 2280
Db 2221 TTCTAAGCAAACTAACTTACATCAAAATAATTTTTTTTGTGAAAGATTTAAACAAAT 2280
Qy 2281 AGTTGAAAAATTTAATGAGGATCAAGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 2340
Db 2281 AGTTGAAAAATTTAATGAGGATCAAGAAATGAAATGAAATGAAATGAAATGAAATGAAAT 2340
Qy 2341 TGGCTTATCAATAATTTTCTCAGATTTTACCTGGAGAGGATACAGATATTTATCATCTTCT 2400
Db 2341 TGGCTTATCAATAATTTTCTCAGATTTTACCTGGAGAGGATACAGATATTTATCATCTTCT 2400

Qy 2401 CTCAGAGAAATGCACACCAATTAGTTCTCTCAGAAATACCTGTATAAATTTATTTAAATATTC 2460
Db 2401 CTCAGAGAAATGCACACCAATTAGTTCTCTCAGAAATACCTGTATAAATTTATTTAAATATTC 2460
Qy 2461 TCAGGATTTATATGTTGAAATTTACAAAAGATGAGCAAAAATATAAAGAAAAATAGGATATA 2520
Db 2461 TCAGGATTTATATGTTGAAATTTACAAAAGATGAGCAAAAATATAAAGAAAAATAGGATATA 2520
Qy 2521 TGAACGAGTTTAAATGTTTACAGATTTATTTCTTAATATATCAGAAAAAATCTATTGATAATGT 2580
Db 2521 TGAACGAGTTTAAATGTTTACAGATTTATTTCTTAATATATCAGAAAAAATCTATTGATAATGT 2580
Qy 2581 ACTGTTTGAATTTTATTAAGAATGTATCGAGCTTTTGAATCTATTTTACAAAGATTTGTT 2640
Db 2581 ACTGTTTGAATTTTATTAAGAATGTATCGAGCTTTTGAATCTATTTTACAAAGATTTGTT 2640
Qy 2641 GTTTATTTGATAGAAATAAAAAATGTTCTAAGAAATGAAGTTTGGTCTTAATTTGGTTTCG 2700
Db 2641 GTTTATTTGATAGAAATAAAAAATGTTCTAAGAAATGAAGTTTGGTCTTAATTTGGTTTCG 2700
Qy 2701 CTTTCCATGATTTTGTGGCAATTTCTTTTATCAAAATGAAAAACGAAAAACAGCTTATTTATTT 2760
Db 2701 CTTTCCATGATTTTGTGGCAATTTCTTTTATCAAAATGAAAAACGAAAAACAGCTTATTTATTT 2760
Qy 2761 AAGTAATCTAAATGTCAGATGAACTATTTTATACAGACAATTTATAGAAAAATATGAATTT 2820
Db 2761 AAGTAATCTAAATGTCAGATGAACTATTTTATACAGACAATTTATAGAAAAATATGAATTT 2820
Qy 2821 TCAAAATAGATTTCTAAATATGAAATTTAAGATATATAAAGTGGAAAAAATCAACATCT 2880
Db 2821 TCAAAATAGATTTCTAAATATGAAATTTAAGATATATAAAGTGGAAAAAATCAACATCT 2880
Qy 2881 TCTCCTATTGCTTTACAGATGATTTCTATTGATGAATTTGCTAAATGCAAGAAATTTTAGGT 2940
Db 2881 TCTCCTATTGCTTTACAGATGATTTCTATTGATGAATTTGCTAAATGCAAGAAATTTTAGGT 2940
Qy 2941 TTTTATTTGCTAGAAAGTTAAAAATAGAAAAATFAAATCTAAAAATTTAAAGAAAAATTTACT 3000
Db 2941 TTTTATTTGCTAGAAAGTTAAAAATAGAAAAATFAAATCTAAAAATTTAAAGAAAAATTTACT 3000
Qy 3001 AAAAAATAAAAATAGTTGATTTTCTGAGAGTAAGTATGTTTAAATTTAAATATTTAAATATCACC 3060
Db 3001 AAAAAATAAAAATAGTTGATTTTCTGAGAGTAAGTATGTTTAAATTTAAATATTTAAATATCACC 3060
Qy 3061 CGGAATATTTTATTTTAACTTCTGTTGAGAGTAAGTATGTTTAAATTTAAATATTTAAATATG 3120
Db 3061 CGGAATATTTTATTTTAACTTCTGTTGAGAGTAAGTATGTTTAAATTTAAATATTTAAATATG 3120
Qy 3121 TATTTTATTTAATTTTATGAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 3180
Db 3121 TATTTTATTTAATTTTATGAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 3180
Qy 3181 TAATATTTAAAAATGAAATTTTATGTTTATTTATGTTCTATATATATGTTTGTGTTTCTAG 3240
Db 3181 TAATATTTAAAAATGAAATTTTATGTTTATTTATGTTCTATATATATGTTTGTGTTTCTAG 3240
Qy 3241 TAGTCACAGTATGTTTGTGAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 3300
Db 3241 TAGTCACAGTATGTTTGTGAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 3300
Qy 3301 CCATAATTTGGATTTTGCATTAATGATTAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 3360
Db 3301 CCATAATTTGGATTTTGCATTAATGATTAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 3360
Qy 3361 ATAAAAATTTAAAAATAGTATCTTTTATGTTTATTTAGTTTATTTAGTTTATTTAGTTTATTTAGTT 3420
Db 3361 ATAAAAATTTAAAAATAGTATCTTTTATGTTTATTTAGTTTATTTAGTTTATTTAGTTTATTTAGTT 3420
Qy 3421 TGTATATTTTCAAAATGGGAAAGATATTTGATTTTTTTTAGACAGACACCTTTATAGGACTAG 3480
Db 3421 TGTATATTTTCAAAATGGGAAAGATATTTGATTTTTTTTAGACAGACACCTTTATAGGACTAG 3480

| | | | |
|----|------|------------------------------------------------------------------|------|
| Qy | 3481 | ACTATCTTATAACAGCGGTCAAAAACAGAGTTGGTTGGCTTTATGAACTATCTACGTTAA | 3540 |
| Db | 3481 | ACTATCTTATAACAGCGGTCAAAAACAGAGTTGGTTGGCTTTATGAACTATCTACGTTAA | 3540 |
| Qy | 3541 | ATACCACTACAAATATAGTTTCAATTCGGTTAACTCTTTGCACTTATAAAAAATAAAATGC | 3600 |
| Db | 3541 | ATACCACTACAAATATAGTTTCAATTCGGTTAACTCTTTGCACTTATAAAAAATAAAATGC | 3600 |
| Qy | 3601 | AACAATTTTTTTCTTGCTCTTGCTTTTATACCGATCTATTTAAGTGGATCGAGAATTG | 3660 |
| Db | 3601 | AACAATTTTTTTCTTGCTCTTGCTTTTATACCGATCTATTTAAGTGGATCGAGAATTG | 3660 |
| Qy | 3661 | GTAGTTTATCCGTACCAATATTAATTAATGCTTTGTTATGAGATATATAGTGGGAAAT | 3720 |
| Db | 3661 | GTAGTTTATCCGTACCAATATTAATTAATGCTTTGTTATGAGATATATAGTGGGAAAT | 3720 |
| Qy | 3721 | TTGCTTGGATAAAAAGCTAATAGTAATATTTGTAATACTACTTATTTTAAATACTG | 3780 |
| Db | 3721 | TTGCTTGGATAAAAAGCTAATAGTAATATTTGTAATACTACTTATTTTAAATACTG | 3780 |
| Qy | 3781 | AATTTGCTTTTACCATGAAATTTTGGCTGTTTATAATTTCTAGAGAATCAAGTAACGAAGCTA | 3840 |
| Db | 3781 | AATTTGCTTTTACCATGAAATTTTGGCTGTTTATAATTTCTAGAGAATCAAGTAACGAAGCTA | 3840 |
| Qy | 3841 | GATTTATATTTTATCAAGGAAGTATTTGATAAAGTATTAGAAAACAATATTTTATTTGGAT | 3900 |
| Db | 3841 | GATTTATATTTTATCAAGGAAGTATTTGATAAAGTATTAGAAAACAATATTTTATTTGGAT | 3900 |
| Qy | 3901 | ATGGAATATCCGAATATTCAGTTACGGGAATTCGGCTCGGAAGTCATTCAGGCTATATAT | 3960 |
| Db | 3901 | ATGGAATATCCGAATATTCAGTTACGGGAATTCGGCTCGGAAGTCATTCAGGCTATATAT | 3960 |
| Qy | 3961 | CAATTTTTTATAAATCAGGAATAGTTGGTTGATTTTACTGATGTTTCTTTTTTTATG | 4020 |
| Db | 3961 | CAATTTTTTATAAATCAGGAATAGTTGGTTGATTTTACTGATGTTTCTTTTTTTATG | 4020 |
| Qy | 4021 | TTATAAAAAAAGTTATGGAGTTAATGGGGAACAGCACTATTTTATTTTACATCATTAG | 4080 |
| Db | 4021 | TTATAAAAAAAGTTATGGAGTTAATGGGGAACAGCACTATTTTATTTTACATCATTAG | 4080 |
| Qy | 4081 | CCATATTTTTCATATATGAACAATAGATCCGATTTATTTATTTATTTAGTACTATCTTTT | 4140 |
| Db | 4081 | CCATATTTTTCATATATGAACAATAGATCCGATTTATTTATTTATTTAGTACTATCTTTT | 4140 |
| Qy | 4141 | CTTCAATAGGTATTTGGAATATTAATTTTAAAGAGGATATCGAGCAAAAAATGAAT | 4200 |
| Db | 4141 | CTTCAATAGGTATTTGGAATATTAATTTTAAAGAGGATATCGAGCAAAAAATGAAT | 4200 |
| Qy | 4201 | GATTTAATTTTCAAGTTATTTGTAACCAATTTATTAATGTCCAAGATTTATCTTGATAAATGAT | 4260 |
| Db | 4201 | GATTTAATTTTCAAGTTATTTGTAACCAATTTATTAATGTCCAAGATTTATCTTGATAAATGAT | 4260 |
| Qy | 4261 | AACAGTATTTTACCAAAACATATCTAATTTAGAGGTTATCTCGTAATATGATGGAGT | 4320 |
| Db | 4261 | AACAGTATTTTACCAAAACATATCTAATTTAGAGGTTATCTCGTAATATGATGGAGT | 4320 |
| Qy | 4321 | ACTGATGTTCTGAGAAAATTTGCTTAACTATATGAAGACGATGGAAGAAATTAATAT | 4380 |
| Db | 4321 | ACTGATGTTCTGAGAAAATTTGCTTAACTATATGAAGACGATGGAAGAAATTAATAT | 4380 |
| Qy | 4381 | TACAAGAAAATTAATGCGGTCTAGCAGATGCTCGAAATTTTCGGACTAGAACATGCAACA | 4440 |
| Db | 4381 | TACAAGAAAATTAATGCGGTCTAGCAGATGCTCGAAATTTTCGGACTAGAACATGCAACA | 4440 |
| Qy | 4441 | GGTAAATATATGCTTTTGTGCTGATCACTATATAGAAGTTGCAATGTTTCGAGAGA | 4500 |
| Db | 4441 | GGTAAATATATGCTTTTGTGCTGATCACTATATAGAAGTTGCAATGTTTCGAGAGA | 4500 |
| Qy | 4501 | ATGCATGATATATACCTGAGTATAATGCCCATATATAGCAGATAGATTTTGTGTAGTA | 4560 |
| Db | 4501 | ATGCATGATATATACCTGAGTATAATGCCCATATATAGCAGATAGATTTTGTGTAGTA | 4560 |
| Qy | 4561 | GACGAAAACGGGTATACAAAGAAAAGAAATAGTAATTTTCATGTTCTTAACGAGAGA | 4620 |

| | | | |
|----|------|-----------------------------------------------------------------|------|
| Db | 4561 | GACGAAAACGGGTATACAAAGAAAAGAAATAGTAATTTTCATGTTCTTAACGAGAGA | 4620 |
| Qy | 4621 | GAGACTGTAAAGAAATTTTGTTCAGGATCTAATATAGAAAATAATTTTGTGTCGAAGCTT | 4680 |
| Db | 4621 | GAGACTGTAAAGAAATTTTGTTCAGGATCTAATATAGAAAATAATTTTGTGTCGAAGCTT | 4680 |
| Qy | 4681 | TATTCAGCAGATATTAATAAGATATAAATTTCCAAATTTAATAATAGAAATTTTGGTGAG | 4740 |
| Db | 4681 | TATTCAGCAGATATTAATAAGATATAAATTTCCAAATTTAATAATAGAAATTTTGGTGAG | 4740 |
| Qy | 4741 | GATTTGCTTTTAAATTTGGAGGCTTCGAACAATGTAAACACGTGTAGTTGATCTAGA | 4800 |
| Db | 4741 | GATTTGCTTTTAAATTTGGAGGCTTCGAACAATGTAAACACGTGTAGTTGATCTAGA | 4800 |
| Qy | 4801 | GAATATTTATTAATTAATGTCAATTCGTAAACAGTTTCGCTTATTAATTCAGAAATTTCTATA | 4860 |
| Db | 4801 | GAATATTTATTAATTAATGTCAATTCGTAAACAGTTTCGCTTATTAATTCAGAAATTTCTATA | 4860 |
| Qy | 4861 | AATAATATGATTTAGTCAAGATTCGAGAAATTTACCCCTTTAAAGTTAAAAAGAGAGTTT | 4920 |
| Db | 4861 | AATAATATGATTTAGTCAAGATTCGAGAAATTTACCCCTTTAAAGTTAAAAAGAGAGTTT | 4920 |
| Qy | 4921 | AGTCATTTATTTGATGCAAAAAGTTTATTAAGAGAGGTTTAAATGTTTAAAACAAAATGTAT | 4980 |
| Db | 4921 | AGTCATTTATTTGATGCAAAAAGTTTATTAAGAGAGGTTTAAATGTTTAAAACAAAATGTAT | 4980 |
| Qy | 4981 | TCAACAGATTTGTTGGATTAATGAGTCTTCGCCAATATTAGAGTCTTATCGAAAAAGAAATA | 5040 |
| Db | 4981 | TCAACAGATTTGTTGGATTAATGAGTCTTCGCCAATATTAGAGTCTTATCGAAAAAGAAATA | 5040 |
| Qy | 5041 | CGTAGATATCCATTTTAAAGCGAAAAGATATTATCAAGAAAGCAATTTAGTTACGTTG | 5100 |
| Db | 5041 | CGTAGATATCCATTTTAAAGCGAAAAGATATTATCAAGAAAGCAATTTAGTTACGTTG | 5100 |
| Qy | 5101 | TATTTGATGAAATTTTCGCCCTAAACTATATATGTAATGTTTATATAGAAATTTCAAAAGCAG | 5160 |
| Db | 5101 | TATTTGATGAAATTTTCGCCCTAAACTATATATGTAATGTTTATATAGAAATTTCAAAAGCAG | 5160 |
| Qy | 5161 | TAGAGGTAATAAATGGAATAAATTAGTGTATTGTTCCAGTTTATAAAGTACATAAATTT | 5220 |
| Db | 5161 | TAGAGGTAATAAATGGAATAAATTAGTGTATTGTTCCAGTTTATAAAGTACATAAATTT | 5220 |
| Qy | 5221 | TAAGTAGTTGTATAGAAAGCAATTTAATCAAAATTTAAAAATATAGAAATATATTGA | 5280 |
| Db | 5221 | TAAGTAGTTGTATAGAAAGCAATTTAATCAAAATTTAAAAATATAGAAATATATTGA | 5280 |
| Qy | 5281 | TAGATGATGGCTCTGTAGATGATTTCTGCTAAATATGCAAGGAATATGCAAAAAAGATA | 5340 |
| Db | 5281 | TAGATGATGGCTCTGTAGATGATTTCTGCTAAATATGCAAGGAATATGCAAAAAAGATA | 5340 |
| Qy | 5341 | AAAGAGTAAATAATTTTTCACATACTAGTGGAGTATCAAAATGCTAGAAATCATGCAA | 5400 |
| Db | 5341 | AAAGAGTAAATAATTTTTCACATACTAGTGGAGTATCAAAATGCTAGAAATCATGCAA | 5400 |
| Qy | 5401 | TAAAGCGGAGTACAGCTGGAATATTTATGTTGTTGCTCTGATGATGTTGTTGATAGTA | 5460 |
| Db | 5401 | TAAAGCGGAGTACAGCTGGAATATTTATGTTGTTGCTCTGATGATGTTGTTGATAGTA | 5460 |
| Qy | 5461 | GATTTAGTAAAAAATTTTATTTAATTTTAAAAAGTAGAAGTGAATTTCTCGTTGTT | 5520 |
| Db | 5461 | GATTTAGTAAAAAATTTTATTTAATTTTAAAAAGTAGAAGTGAATTTCTCGTTGTT | 5520 |
| Qy | 5521 | TGTAAGCTACTTTTTCAGAAAATTAATAATTTTGAAGTGAATAATTTCAAAATATTGATT | 5580 |
| Db | 5521 | TGTAAGCTACTTTTTCAGAAAATTAATAATTTTGAAGTGAATAATTTCAAAATATTGATT | 5580 |
| Qy | 5581 | TTGAAGCAATTTAATACCGTCAGACATGCGAGAAAATAATTTTGAATTTGATATAA | 5640 |
| Db | 5581 | TTGAAGCAATTTAATACCGTCAGACATGCGAGAAAATAATTTTGAATTTGATATAA | 5640 |
| Qy | 5641 | ATATATATTTTCTTACTCTCTGTTTGTAAACTATATATAAGAAAAGATACATAACAGATCTTT | 5700 |


```
US-10-865-873-3
; Sequence 3, Application US/10865873
; Publication No. US20050064559A1
; GENERAL INFORMATION:
; APPLICANT: Miyake, Katsuhide
; APPLICANT: Watanabe, Masaki
; APPLICANT: Iijima, Shinji
; TITLE OF INVENTION: Beta 1,3-galactosyltransferase and DNA encoding the same
; FILE REFERENCE: 786.53
; CURRENT APPLICATION NUMBER: US/10/865,873
; CURRENT FILING DATE: 2004-06-14
; PRIOR FILING DATE: US/09/900,038
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: JP 2001-392
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6865
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae type Ib
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (617)..(1789)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1816)..(2262)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2265)..(2744)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2843)..(3979)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3982)..(4953)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5009)..(5947)
; US-10-865-873-3

Query Match 10.4%; Score 727.8; DB 19; Length 6865;
Best Local Similarity 59.2%; Pred. No. 9.1e-93;
Matches 1389; Conservative 0; Mismatches 917; Indels 41; Gaps 7;

Qy 16 TGGCAATATTGATGATAGCAGTTGCAATTTCTGCAATCTTAAACAAGTCATATACCAA 75
Db 438 TAGCGATGATTCAAACCGTTGTGGTTTATTTCTGCAAGTTTGACATTAACTAATTA 497

Qy 76 ATGCTGATTTAAATCG---TTCTCGAATTTTATCATATGATGTTTCATTTTTCAT 132
Db 498 CTCCCAACTTTAAAGCAATAAAGATTATTGTTGTTCTATTGATACATTATATTGCT 557

Qy 133 TTTTATATCTCGATGCGAGTTGCAATTTTGAGTATAGAGGTAATCTGATAGAGTTTCAAA 192
Db 558 TTTATCTTTCTGATTTTACAGAGCTTTTGGAGTCTGGTCTATCTTGAAGAGTTTAAAA 617

Qy 193 AAACATTAACCTATAGTATATATTTGCAATTTTCTTACGGCAGTATCATTTTGTGG 252
Db 618 TGGTATTGAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTATTTTATTTCTA 677

Qy 253 AGAATAATTTCCGACTTTCAAGAGCTGGTGGCCGTGATTTTCATATTAATAACTTCGTTT 312
Db 678 AAAACTCTTTTACACGACAGCTTTCTCTTTTACTTTTCTTTTATTTGCTATGAAATTCGATT 737

Qy 313 TGGTATACCTATTTAACTTAATTAAGCAGTTTAAAGGATAGCTTTCTATTTTTCGACAA 372
Db 738 TATTATATCTATGAAATTCATTTTAAATATATATTCGAAATATCTTACGCTAAGTTT 797

Qy 373 TCTATCAAAAAGACCGATTTCTAATTAACAACGGCTGAACGATGGGAAAAATATGCAAGTTT 432
Db 798 CACGAGATACCAAAGTTGTTTGTATACGAATAAGGATTTCTTTATCAAAAATGACCTTTA 857

433 TATTTGAATCACATAAACAATTTCAAAAAATCTTGTTCATTTGATTTTGTAGGTACAG 492
858 GGAATAAAT---ACGACCATAAATTATATCGCTGCTGTGATCTTTGGATTTCTCTGAAAAGG 914
493 AAATAGATAAAAAATTAATTTATCATTTACCGCTCTATTATTCTGTGGGAAGACTATAGAGT 552
915 ATTGTTATGATTTGAAACAATACTCGTTAAGGATAATAAACAAGATGCTCTTACTTTCAG 974
553 TTTCAACAAGGGAAGTGGTCGACCAAGCTCTTTATAAATCTACCAAGTGAAGTTTGTAGACG 612
975 AGTTAACTGCTTAACTGTTGATCAAGCTTTTATTAAATACATACCAATTAATTTTGGTA 1034
613 TAAAGCAATTC-----GTTTCAGATTTTCAGTTGTTAGGTATTGATGTAAGCGTTG 663
1035 AATACCAAAATACAAGATATTATTATGACATTCAAGCAATGGGAGTGAATGTCAATGTTA 1094
664 ATATTAAATCATTCGGTTTACTGCGTTGAAAAAACAATAAATCCAACTCTAGGTGAC 723
1095 ATGTAGAGGCACTTAGCTTTGATAATATAGGAGAAAGCGAATCCAACTTTTGAAGGAT 1154
724 ATAGCATTTGTAATTTTTCACAAAATTTTATAAGCCTAGTCTATATCATGATGAAGAC 783
1155 ATAGTGTATTACATATTCTATGAAATTTCTATAAATATAGTCACTTATAGCAAAACGAT 1214
784 TTTTGGATATACCTCGAGCGGTAGTCGGGTTAATTTTGTGTTAGTATTCTTATTTGT 843
1215 TTTTGGATATCATGGGTGCTATTATAGGTTTGTCTATATGTGGCATTTGTGGCAATTTTTC 1274
844 TAGTTCCAATTTATTCGTAGAGATGGTGGACCGGCTATTTTGTCTAGAAAACGAGTTTGGAC 903
1275 TAGTTCCGCAATCAGAAAAGATGGTGGACCGGCTATCTTTTCTCAAAATAGATAGTCTC 1334
904 AGAATGAGCGCATTTTACATTTCAAGTTTTCGATCGATGATGATGTTGATGCTGAGAGC 963
1335 GTAAATGTTAGGATTTTGTAGATTCTATAAATTCAGATCAATGCGAGTAGATGCGAACA 1394
964 GCAAAAAGACTTGTCTCAGCCAAAACAGATGCAAGGTGGGTATGTTTAAATGGAA 1023
1395 TTAAGAAAGATTTATTAGTTTCAANTCA-----AATGACGGGCTAATGTTTAAAGTTAGA 1449
1024 AAACGATCTTAGAATTTACTCCAATTGGACATTTTCATACGCAAAAAACAAGTTTAGACGAT 1083
1450 CGATGATCTTAGAATTTACTAAAATAGAAAATTTTATTCG-AAAAACAAGCATAGATAGT 1508
1084 TACCACAGTTTATAATGTTTAAATTTGGCGATATGATCTAGTTGGTACACGCTCACCTA 1143
1509 TGCCTCAATTTCTATAATGTTTAAAAGGTGATATGATTTAGTAGGAACAACGCGCTCCCA 1568
1144 CAGTTGATGAAATTTGAAAAATATACCTCTGGTCAAAAGAGACGATTCAGTTTAAACAG 1203
1569 CAGTTGATGAAATATGAAAAAGTATAATTCAACGCAAGAGCGACGCTTAGTTTAAAGCAG 1628
1204 GGAATACAGCTCTCTGGCAGGTTAGTGGTCTAGTAAATATCACAGACTTCGACGACGTAG 1263
1629 GAATCACTGGTTTGTGCAAAATATCTGTAGAAATATAATTAATCTGATTTTGTAGAAATCG 1688
1264 TCCGTTGGACTTTAGCATACATTTGATTAATTTGGAATCTATCTGGTCAGATATTAAATTTAT 1323
1689 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTGTCAGATATTAAAGATTATTC 1748
1324 TAAAGCAGTGAAGTTGTTATTTGTCAGAGAGGGAAGTAAAGTA----- 1372
1749 TCCTAACGCTTAAAGGTTAGTTTACTCGGGAAGGAGCTAAGTAAAGGTAAGGTTTGAAG 1808
1373 -----TATGAAAGTTTGTGTTGGTTCGTTCTTCCAGGGGACATTTTGACTCACTTGTATTT 1426
1809 GAATATAATGAAATTTTGTCTGGTTGGTTCAAGTGGTGGTCCCTAGCACACTTGAACCT 1868
1427 GTTAAACCGTTTGGAGGAAGAAAGAAAGCTTTTGGGTTAAACATTTGATTAAGAGGATGC 1486
1869 TTTGAAACCCATTTGGGAAAAAAGAAAGATAGTTTGGGTAACTTTTGTGATTAAGAGATGC 1928
1487 AAGAAGTCTTTTGAAGAAATGAAAAAATGTATCCATGTTACTTTTCCAAACAAATGCAATCT 1546
```


Db 5602 TAGTTCGCAATCAGAAAGATGGTGGACGGCTATCTTTTCTCAAAATAGAGTAGGTC 5661
Qy 904 AGAATGACGCAATATTTACATCTTACAAAGTTTCATCGATGTATGTTGATGCTGAGGAGC 963
Db 5662 GTAATGGTAGGATTTTATAGATTTCTATATAATTCAGATCAATGCGAGTAGATGCGAAACAAA 5721
Qy 964 GCAAAAAGACTTGCTCAGCCCAACCAACAGATGCAAGGGTGGGTATGTTTAAATGGAA 1023
Db 5722 TTAAGAAGATTTATTTAGTTGACAAATCAAATGACAGGG-----CTAATGTTAAGTTAGA 5776
Qy 1024 AAACGATCCTAGATTTACTCCAAATGGACATTTTCATACGCAAAACAAAGTTTACAGAGT 1083
Db 5777 AGATGATCCTAGATTTACTAAATAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT 5835
Qy 1084 TACCACAGTTTATAATGTTTAAATGCGGATATGAGTCTAGTGTGGTACACGCTCCACTA 1143
Db 5836 TGCCTCAATTTCTAATATGTTTAAAGGCGATATGAGTTTAGCAGGAACACGCGCTCCCA 5895
Qy 1144 CAGTTGATGAATTTGAAAATATATCTCTGGTCAAAAGACGATTCAGTTTAAACCCAG 1203
Db 5896 CAGTTGATGAATATGAAAAGTATTAATTCACGCAAGACGACGCGCTTAGTTTAAAGCCAG 5955
Qy 1204 GGATTACAGGTCTCTGCGCAGGTTAGTGGTCTAGTATATACACAGCTTCGACGACGTAG 1263
Db 5956 GAATCACTGGTTTGTGCAAAATATCTGTAGAAATATATATCTGATTTTGATGAATCG 6015
Qy 1264 TTCGGTTGGACTTAGCATACATTTGATTAATTTGGACTCTCTGGTACAGATATTTAAATTTAT 1323
Db 6016 TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTTGGTACAGATTAATGAATTTATC 6075
Qy 1324 TAAAGACAGTGAAGTTGTTATTTGTTGAGAGGGAAGTAAGTAAAGTA----- 1372
Db 6076 TCCTAAACACTAAAGGTAGTTTACTCGGACAGAGCTAAGTAAAGTAAAGTTTGAAG 6135
Qy 1373 -----TATGAAAGTTGTTTGGTGGTCTTTCAGGGGACATTTGACTCACTTGATTT 1426
Db 6136 GAATATATGAAATTTGTCGGTTGGTTCAAGTGGTGGTCACTAGCACACTTGAACCT 6195
Qy 1427 GTTAAACCGTTTGGAGGAGGAAGAACGTTTGGGTAAACATTTTGATAAAGGAGTGC 1486
Db 6196 TTTGAAACCCATTTGGGAAAGGAAGATAGGTTTGGGTAAACCTTTGATAAAGAGATGC 6255
Qy 1487 AAGAGTCTTTTGAAGATGAAAATGTATCCATGTACTTCCACAAATCGCATCT 1546
Db 6256 TAGGATATTTCTAAGAGAGAGATTGTATATCATTTGCTTTCTTCCAAACCAACCGTAAATGT 6315
Qy 1547 CAITTAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACGAGA 1606
Db 6316 CAAAAACTTGGTAAATAATFACTATTCTAGCTTTTAAAGTCTTTAGAAAGAGAACCCAGA 6375
Qy 1607 TGTTATTTATTCATCTGGTGGCGGTTGCTGTCCTTCTTTTACATCGGAAACATTTAT 1666
Db 6376 TGTTATCATATCATCTGGTGGCGCTGAGCAGTACCATTCTTTTATATTTGTTAAGTTATT 6435
Qy 1667 TGGAGCAAGACGATTTATTTAGAGTATTTGATCGAGTTAATTAATCTACATTAACG 1726
Db 6436 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACCACTTTGACAG 6495
Qy 1727 AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCAAGGGAAGAAATGAAGAGT 1786
Db 6496 AAAATTAGTGTATCTGTAAACAGATAAATTTATTTGTTTCAAGTGGGAAGAAATGAAGAGT 6555
Qy 1787 ATATCTTAATCTATTAACCTGGGAGTATTTTAAATGATTTTGTAAAGAGTGGAGT 1846
Db 6556 TTATCCTTAAGCAATTAATTTAGGAGGAAATTTTAAATGATTTTGTACAGTGGGACA 6615
Qy 1847 CATGAACACAGTTTAAATCGATTGATAAAGAGATTTGATTTTGAAGAGTGGAGT 1906
Db 6616 CATGAACAGAGTTTCAACCGCTTTATTAAGAGAGTTGATAGATTTAAAGGGACAGGTGCT 6675
Qy 1907 ATAAACGACGAAATATTTATTTCAACAGAGGATTTCTGACTATATTTCCAGAAATTTGCAAG 1966
Db 6676 ATTGATCAAGAGTGTTCATTTCAACCGGGTTACTCAGACTTCGAACTTCAGAAATTTGTCAG 6735

Qy 1967 TATAAAAAATTTCTCAGTTTACAAAGAAATGGAACAATATATTAACAACATCAGAAGTAGTT 2026
Db 6736 TGGTCAAAAATTTCTCTCATATGATGATGAACCTTTACATGAAGAAGCTGAGATTTGTT 6795
Qy 2027 ATTTCGACGAGGCCCGCTACTTTTATGAATTCATTTATCCAAAGGAAAAACAATTA 2086
Db 6796 ATCACACATGGCGGCCAGCGGTTTATGTCTAGTTATTTCTTTAGGGAATTTACCAAGTT 6855
Qy 2087 TTGTTTCTTAGACAAAAAAGTATGCTGAACATGTAAATGATCATCAAGTAGAGTTTGTGA 2146
Db 6856 GTTGTTCCTTAGGAAAGCAGTTTGGTGAACATATCAATGATCATCAAAATACAATTTTGA 6915
Qy 2147 AGAAGAAATTTTACAAGATAAATAATTTATTTATAGAAAATATAGATGATTTGTTGAA 2206
Db 6916 AAAAAAATTTGCCACCTGTATCTTGGCTTGGATTTGAAGATGTAGATGCACTTTGCGAA 6975
Qy 2207 AAAATTTATGAAGTTTCTAAGCAAACTAACTTTTACATCAATAATAATTTTGTGAA 2266
Db 6976 GCCTTGAAGGAATATAGCTACAGAAAATATCAGGGAATTAATGATATGTTTGTGTCAT 7035
Qy 2267 AGATTAACAAACATAGTTTGAATAAT 2292
Db 7036 AAATTAGAAAAAATTATAGTGAAAT 7061

RESULT 6

US-09-751-708A-83
; Sequence 83, Application US/09751708A
; Publication No. US20030157113A1
; GENERAL INFORMATION:
; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 751708
; CURRENT APPLICATION NUMBER: US/09/751,708A
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 60/173,371
; PRIOR FILING DATE: 1999-12-28
; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 83
; LENGTH: 17276
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7062)..(8207)
; OTHER INFORMATION:
US-09-751-708A-83

Query Match 10.4%; Score 726.8; DB 10; Length 17276;
Best Local Similarity 59.2%; Pred. No. 1.7e-92;
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

Qy 16 TGGCATATTTGATATGATGAGCTGGCAATTTCTGCAATCTTAAACAAGTCATATACCAA 75
Db 4765 TAGCGATGATTCAAACAGTTGTAGTTATTTTCTGCAAGTTTGACATTAACATTAATTA 4824
Qy 76 ATGCTCATTTAATCG---TTCCTGGAATTTTATCATATGATGGTTTCATTATTTGTCAT 132
Db 4825 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTCTTATGATACATATATTTGTTT 4884
Qy 133 TTTTATATCTCGTATGCCAGTTTGAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAAA 192
Db 4885 TTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 4944
Qy 193 AAACATTTAATATAGTATTAATTTTGGCAATTTTCTTACGGCAGTATCATTTTGTGG 252
Db 4945 TGGTATTTGAATAACAGCTTTTACTATATTTTCAATATCAAGTTTCATTATTTTATTTTA 5004
Qy 253 AGAATAATTTGCGCACTTCAAGACGGTGGCGGTGATTTTACATTAATAAATTCGTTT 312
Db 5005 AAAACTCTTTTACACGACAGCACTTTCCTTTTACTTTTATTTGCTATGAATTCGATTT 5064


```
; CURRENT APPLICATION NUMBER: US/10/428,817A
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US 60/378,988
; PRIOR FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: US 60/389,366
; PRIOR FILING DATE: 2002-06-15
; PRIOR APPLICATION NUMBER: US 60/406,697
; PRIOR FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: US 60/406,750
; PRIOR FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: US 60/415,310
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 60/415,400
; PRIOR FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: US 60/438,686
; PRIOR FILING DATE: 2003-01-09
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 79
; LENGTH: 17276
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7062)..(8207)
US-10-428-817A-79

Query Match      10.4%; Score 726.8; DB 18; Length 17276;
Best Local Similarity 59.2%; Pred. No. 1.7e-92;
Matches 1366; Conservative 0; Mismatches 902; Indels 38; Gaps 6;

Qy      16  TGGCAATATTGATATGATAGCAGTTCGAATTTCTGCAATCTTAAACAGTCATATACCAA 75
Db      4765 TAGCGATGATTCAAACAGTTGTAGTTTATTTCTGCAAGTTTGACATTAACATTAATTA 4824

Qy      76  ATGTCGATTTAAATCG--TTCTGGAATTTTATCATAATGATGGTTTCATTTATTTTCAT 132
Db      4825 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATTAATTTT 4884

Qy      133 TTTTATATCTCGATGCGAGTTGAAATTTGAGTATAGAGTAAATCTGATAGAGTTTGA 192
Db      4885 TTTATCTTCTGATTTTACAGACACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAA 4944

Qy      193 AAACATTTAACTATAGTATATATTTGCAATTTTCTTACGCGAGTATCATTTTGTGG 252
Db      4945 TGGTATTGAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTATTTTATTT 5004

Qy      253 AGAATAATTTTCGCACTTTCAAGACGTGGTCCGCTGTTTTCACATTAATAAATTCGTT 312
Db      5005 AAACCTCTTTTACAGCAGCACTTTCCCTTTTATCTTTTATTTTATGCTATGAAATTC 5064

Qy      313 TGGTATACCTATTTTAACTATTAATTAAGCAGTTTAAAGGATAGCTTTCTATTTTCGCAA 372
Db      5065 TATTATATCTATTGAAATTCATTTTAAATAATTTATCGAAATAATTTCTTACGCTAAGTT 5124

Qy      373 TCTATCAAAAAGACAGTCTTAATTAACACGGCTGAACGATGGGAAATTAAGCAAGTT 432
Db      5125 CACGAGATACCAAGTTGTTTGTATGATAAGGATTTCTTTATCAAAAATGACCTTTTA 5184

Qy      433 TATTTGAATCATAATAACAAATTCAAAAAATCTGTTGCAATGCTAGTTTATAGGTACAG 492
Db      5185 GGAATAAAT---ACGACCAATTAATATTCGCTGTCTGTATCTTGGACTCCTCTGAAAGG 5241

Qy      493 AAATAGATAAAATTAATTTATCATTCACGCTCTATTATTTCTGTGGAAAGAGCTATAGAGT 552
Db      5242 ATTGTTATGATTTGAAACATAAATCGTTAAGGATAATAAACAAGATGCTCTTACTTT 5301

Qy      553 TTTCACAGGGAAGTGTGACACAGCTCTTTATAATCTTACAAAGTGAGTTTATAGACG 612
Db      5302 AGTTAACTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCAATTTGAATTTATTTGGTA 5361

Qy      613 TAAAGCAA-----TTCTGTTTCAGATTTTGAAGTTGTTAGGTATTTGATTAAGCGCTTG 663
Db
```

```
Db      5362 AATACCAATACAGATATTTAATGACATTTGAAGCAATGGGAGTGATTTGTCAATGTTA 5421
Qy      664 ATATTAAATCAATTCGGTTCGTTTACCTGCGTTGAAAAACAAAAATCCAACTCCTAGTGCACC 723
Db      5422 ATGTAGAGGCACCTAGCTTTGATAATATAGGAGAAAAAGCGAAATCCAACTTTTGAAGGAT 5481
Qy      724 ATAGCAATTTGAATTTTCCCAAAATTTTATAGCCTAGTCTATATCATCATGATGAACGAC 783
Db      5482 ATAGTGTATTACATATTTCTATGAAAATTTCTATAAATATATAGTCCCTTATAGCAAAACGAT 5541
Qy      784 TTTTGGATATACCTCGGAGCGGTAGTCGGGTTAATTTTGTGTATAGTTTCTTATTTTGT 843
Db      5542 TTTTGGATATACCGGTGCTATTATAGTTTGTCTATATGTCATTTGTGCAATTTGTGCAATTTTTC 5601
Qy      844 TAGTTCCAAATTTCTGAGATGTTGGACCGGCTATTTTGTCTAGAAAACGATTTGGAC 903
Db      5602 TAGTTCGCAAAATCAGAAAAAGATGGTGGACCGGCTATCTTTCTCAAAATAGAGTAGGTGC 5661
Qy      904 AGAATGAGCGCATATTTACATTTCTCAAGTTTTCGATCGATGTATTTGATGCTGAGGAGC 963
Db      5662 GTAATGGTAGGATTTTGTAGATTTCTATAAATTCAGATCAATGCGAGTAGATGCGAAACAA 5721
Qy      964 GCAAAAAGACTTTGCTCAGCCAAAACAGATGCAAGGTGGGTATGTTTAAAAATGGAA 1023
Db      5722 TTAAGGAAGATTTTATAGTTTCAATCAATGACAGGG-----CTAATGTTTAAAGTTAGA 5776
Qy      1024 AAACGATCCTAGAAATTTACTCCAATTTGACATTTTATACGCAAAAACAAAGTTTATAGACGAT 1083
Db      5777 AGATGATCCTAGAAATTTACTAAAATAGCAAAATTTATTTCG-AAAAACAAGCATAGATGAGT 5835
Qy      1084 TACCACAGTTTATATGTTTAAATTTGGCGATATGATCTAGTTGGTACAGCTCCACCTA 1143
Db      5836 TGCCTCAATTTCTATAATGTTTAAAGGCGATATGATTTAGCAGGAACACGCGCTCCCA 5895
Qy      1144 CAGTTGATGAATTTGAAAAATATACCTCGGTCAAAAGACAGCATTTAGTTTAAAAACGAG 1203
Db      5896 CAGTTGATGAATTAAGAAAGTATTAATCAACGCAAGACGACGCTTAGTTTAAAGCAG 5955
Qy      1204 GGAATACAGTCTCTCGCAGGTTAGTGGTGTAGTAAATATACAGACTTCCAGCAGCTAG 1263
Db      5956 GAATCACTGGTTTGTGCAAAATATCTGGTAGAAATAATATTACTGATTTTGTATGAAATCG 6015
Qy      1264 TTCGTTTGGACTTAGCATATGATTAATTTGGACTATCTGGTACAGATATTTAAATTTTAT 1323
Db      6016 TAAAGTTAGATGTTTCAATATATCAATGAAATGGTCTATTTGGTACAGATATTAAGATATTC 6075
Qy      1324 TAAAGACAGTGAAGTTGTTATTTGTCAGAGAGGGAAGTAAGTAAAGTA----- 1372
Db      6076 TCCTAACACTAAAGGTAGTTTACTCGGACAGAGCTAAGTTAAAGGTAAAGTTTGAAG 6135
Qy      1373 -----TATGAAAGTTTGTTCGGTTCGTTTTCAGGGGACATTTGACTCACTTGTATTT 1426
Db      6136 GAATAATAATGAAATTTTGTCTGGTTGGTTCAGTGGTGGTCACTAGCACACTTGAACCT 6195
Qy      1427 GTTAAACCGTTTGGAAAGGAAGAACGTTTTTGGGTAAACATTTCATGATAAAGAGATGC 1486
Db      6196 TTTGAAACCCATTTGGGAAAAAGAAATAGGTGTTTGGGTAACTTTTGTATAAAGAGATGC 6255
Qy      1487 AAGAAAGTCTTTTCAAGAAATGAAAAATGTATCCATGTTTACTTTTCCAAACAAATCGCAATCT 1546
Db      6256 TAGAGATTTCTAAGAGAGAGATTTGTATATCATTTGTTCTTTTCCAAACAAACCGTAATGT 6315
Qy      1547 CATTAATTTAGTGAATAAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACCGA 1606
Db      6316 CAAAAACTTGGTAAAAAATACTATTTCTAGCTTTTAAAGTCTTTAGAAAAAGAAAGACGAGA 6375
Qy      1607 TGTATTATTTCTGTTGGTGGCGCGTTCGTTGCTCCCTTCTTTTATCATCGGAAACTATT 1666
Db      6376 TGTATCATATCATCTGGTGGCGCTGTAGCAGTACCAATCTCTTTTATATGTTAGTATT 6435
Qy      1667 TCGAGCAAAAGACCATTTATATTTGAAGTATTTGATCGAGTTTAAATAATCTCATTTAACTGG 1726
Db      6436 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGATAGATAAAACCAACTTTTGACAGG 6495
```


| | | | | | | |
|----|------|----------|--------------------------------------------------------|----------------------------------------------|--------------------------|------|
| Qy | 1727 | AAAAC | TAGTTTATCCCGTAA | CAGATATTTTATTTGTT | CAGTGGGAAGAAATGAAGAAGGT | 1786 |
| | | | | | | |
| Db | 6496 | AAAAT | TAGTGTATCCTGTAA | CAGATAAATTTATGTT | CAGTGGGAAGAAATGA AAAAAGT | 6555 |
| | | | | | | |
| Qy | 1787 | ATATCCT | TAAATCTATTAACT | TGGGAGTATTTTTTAATGATTTTTTGTAA | CAGTAGGAAC | 1846 |
| | | | | | | |
| Db | 6556 | TTATCCT | TAAAGGCAATTAAT | TTTAGGAGAAATTTTTTAATGATTTTTTGTCA | CAGTGGGACA | 6615 |
| | | | | | | |
| Qy | 1847 | CATGAACA | CAGTTTTAATCGAT | TGATAAAGAGATTCGATTTATTGAAAAAATATGGAAGT | 1906 | |
| | | | | | | |
| Db | 6616 | CATGA | CAGAGTTCACCGTCTTATTTAAGAGAGTTCATAGATTAAAGGGACAGGTGCT | 6675 | | |
| | | | | | | |
| Qy | 1907 | ATAACCG | ACGAATATTTATTTCAAC | ACGATATTTCTGACTATATTCAGATATTTGCAAG | 1966 | |
| | | | | | | |
| Db | 6676 | ATTGAT | CAAGAAGTGTTTCATTTCAA | CGGTTACTCAGACTTCGAACCTCGAATTTGT | CAG | 6735 |
| | | | | | | |
| Qy | 1967 | TATAAAAA | TTTCTCAGTTACAAGAAAT | TGAAACAATATATTAACAATCAGAAGTAGTT | 2026 | |
| | | | | | | |
| Db | 6736 | TGGTCAAA | ATTTCTCTCATATGATGATGAAC | TCTTACATGAAGAAGCTGAGATTTGTT | 6795 | |
| | | | | | | |
| Qy | 2027 | ATTGTCAC | GGAGGCCCGCTACTTTTATGNA | TTCAATTCCTCAAGGAAAAAAACAATTA | 2086 | |
| | | | | | | |
| Db | 6796 | ATCACAT | CGCGGCCCGACGCTTTATGT | CAGTTATTTCTTTAGGGAATTTACCAAGT | 6855 | |
| | | | | | | |
| Qy | 2087 | TTGTTTCT | TAGACAAAAAAGTAGTGGT | GAAATGATCAATCAAGTAGAGTTGTA | 2146 | |
| | | | | | | |
| Db | 6856 | GTTGTTCT | TAGGAGAAAGCAGTTTGGT | GAAATCAATGATCAATCAATACAAATTTT | 6915 | |
| | | | | | | |
| Qy | 2147 | AGAAGAA | TTTTTACAAGATAATAATATTTTATTTATAGAAAAATATAGATGATTTGTTGAA | 2206 | | |
| | | | | | | |
| Db | 6916 | AAAAAAT | TGCCACCTGTATCCCTTGGCTTGGATTCGAGNTGTAGATGGACTTGGGAA | 6975 | | |
| | | | | | | |
| Qy | 2207 | AAAATTA | TGAAGTTTCTAAGCAAACTTA | CACTTTACATAAATAAATTTTTTTTGTGAA | 2266 | |
| | | | | | | |
| Db | 6976 | CGGTTG | AAAAAGGAATA | TAGCTACAGAAAAATATCAGGGAAAAATATGATGTTTGTGTCAT | 7035 | |
| | | | | | | |
| Qy | 2267 | AGATTA | AAAAACAATAGTTTGA AAAAT | 2292 | | |
| | | | | | | |
| Db | 7036 | AAATTTAG | AAAAAATTTATAGTGTGAAT | 7061 | | |
| | | | | | | |

```

RESULT 8
US-10-192-280-1
; Sequence 1, Application US/10192280
; Publication No. US20040009574A1
; GENERAL INFORMATION:
; APPLICANT: Dattagupta, Nanibhushan
; APPLICANT: Shah, Ketan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: STREPTOCOCCUS AGALACTIAE CAPSULAR POLYSACCHARIDE SYNTHESIS
; TITLE OF INVENTION: GENES
; FILE REFERENCE: 475412001300
; CURRENT APPLICATION NUMBER: US/10/192,280
; CURRENT FILING DATE: 2002-07-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 25020
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-192-280-1

```

| | Query Match | 10.3% | Score 721.4; | DB 17; | Length 35020; |
|----|-----------------------|---------------------------------------------------------------|--------------------|-----------------|--------------------|
| | Best Local Similarity | 58.9%; | Pred. No. 1.1e-91; | | |
| | Matches 1389; | Conservative | 0; | Mismatches 926; | Indels 42; Gaps 7; |
| Qy | 16 | TGGCAATTATTGATGATGATAGCAGTTCGCAATTCCTGCAATCTTAAACAGTCATATACCA | 75 | | |
| Db | 5154 | TAGCGATGATTCAAACAGTTGTGGTTTATTTTCTGCAAGTTTGACATTAACATTAAATTA | 5213 | | |
| Qy | 76 | ATGCTGATTAAATTCGTTCTGGAAATTTT-----TATCATATGATGGTTTCATTATTTTGC | 131 | | |

| | | | |
|------|----|---------------------------------------------------------------------|------|
| 5214 | Db | CTCCCAATTTTAAAGCAATAAAGATTATTATGTTGGTCTTAATTTGATACATATATATGTC | 5273 |
| 132 | Qy | TTTTTTATATCTCGTATGCCAGTTGAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAA | 191 |
| 5274 | Db | TTTTATCTTTCTGATTTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGAGTTTAA | 5333 |
| 192 | Qy | AAAACATTTAACTATAGTATAATATTGCAATTTTTCTTACGGCAGTATCATTTTTGTTG | 251 |
| 5334 | Db | ATGGTATTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTTTTTATTTT | 5393 |
| 252 | Qy | GAGAATAAATTTCCGCACTTTCAAGACGCTGGTGCCTGTATTTTCACATTAATAAACTTCGTT | 311 |
| 5394 | Db | AAAACTCATTTACACGACACGACTTCCTTTTTTCTTTATGCTATGAATTCGATTT | 5453 |
| 312 | Qy | TTGGTATACCTATTTAACGTAAATATTAAAGCAAGTTTAAAGGATAGCTTCTATTTCGACA | 371 |
| 5454 | Db | TTATTGTATCTATTGAATTCATTTTTTAAAAATATTATCGAAAAATTCTTACGCTAAGTTT | 5513 |
| 372 | Qy | ATCTATCAAAAAAGACGATTCTAAATTACAAACGGCTGAACGATGGGAAATATGCAAGTT | 431 |
| 5514 | Db | TCAGAGATACCAAAGTTGTTTGTAAACGAATAAGGATTCCTTTATCAAAAAATGACCTTT | 5573 |
| 432 | Qy | TTATTTGAATTCACATAAAACAAATTCAAAAAAATCTTTGTGCATTTGGTAGTTTTAGGTACA | 491 |
| 5574 | Db | AGGAATAAAT---ACGACCAATAATTATATGTCTCTGCATCTTGGACTCCTCTGAAAG | 5630 |
| 492 | Qy | GAAATAGATAAAATTAATTTATCATTCGCTCTATATATTCTGTGGGAAGCTATAGAG | 551 |
| 5631 | Db | GATTTGTTATGATTTGAAACATAACTCGTTAAGGATTAATAAACAAAGATGCTCTTACTTCA | 5690 |
| 552 | Qy | TTTTTCAACAGGGAAGTGGTCGACCAAGTCCTTTATAAATCTACCAAGTGAGTTTTAGAC | 611 |
| 5691 | Db | GAGTTAACCTGCTTAACTGTGTGATCAAGCGTTTATTAAACATACCCATGGAATTTTGGT | 5750 |
| 612 | Qy | GTAAGCAATTCG-----TTTCAGATTTTGGTGTGTAGGTATTTGATGTGAAGCGTT | 662 |
| 5751 | Db | AAATACCAATACAGATATTATTAATGACATTTGAAGCAATGGAGTGATTTGCTCAATGTT | 5810 |
| 663 | Qy | GATATTAATTCATTTCCGTTTTTACTGCGTTTGAAAAACAAAAAATCCAACTGCTAGGTGAC | 722 |
| 5811 | Db | AATCTAGAGCACTTAGCTTTTGATAATATAGGAGAAAAGCAATCCAAACTTTTGAAGGA | 5870 |
| 723 | Qy | CATAGCATTTGTAACTTTTCCACAAATTTTTTATAAGCTTAGTCAATCATATGATGAAACGA | 782 |
| 5871 | Db | TATAGTGTATTACATATTCATCAAAATTCATAAAATATATAGTACCCCTTATAGCAAAACGA | 5930 |
| 783 | Qy | CTTTTGGATATCTCCGAGCGGTAGTCGGGTAAATATTATTGTGSTATAGTTTCTATTTTG | 842 |
| 5931 | Db | TTTTTGGATATCACGGGTGCTATTATAGTTTGTCTCATATGTGGCATTTGGCAATTTTT | 5990 |
| 843 | Qy | TTAGTTCCAAATTTCTGTAGAGATGGTGACCGGCTATTTTTGCTCAGAAACAGTTTGGGA | 902 |
| 5991 | Db | CTAGTTCCCAATCAGAAAGATGGTGGACCGGCTATCTTTTCTCAAAATAGAGTAGGT | 6050 |
| 903 | Qy | CAGAAATGGAACGATATTACATTTCTACAAGTTTCGATCGATGTATGTTGATGCTAGGAG | 962 |
| 6051 | Db | CGTAATGGTAGGATTTTTAGATTCTATAAATTCAGATCAATGGCGAGTAGTCAGAAACAA | 6110 |
| 963 | Qy | CGCAAAAAACATTTGCTCAGCCCAAAACCATGCAAGGGTGGGTATGTTTTTAAAAATGGGA | 1022 |
| 6111 | Db | ATTAAGAAAGATTTATTAGTTTCAATCAAAATGACAGGG-----CTAATGTTTAAGTTAG | 6165 |
| 1023 | Qy | AAAAAGATCTCAGAAATTTACTTCCAAATTTGGAATTTTCATACGCAAAAAACAAAGTTTAGACGAG | 1082 |
| 6166 | Db | ACGATGATCTAGAAATTTACTTAAATAGGAAAAATTTATTCG-AAAAACAGCATAGATGAG | 6224 |
| 1083 | Qy | TTACCACAGTTTTTAAATGTTTTTAAATTTGGCGATATGAGTCTAGTTGGTACACGTCCACCT | 1142 |
| 6225 | Db | TTGCTCAATTCCTATAATGTTTTTAAAAAGGTGATGAGTTTGTAGTAAACACGCGCTCCC | 6284 |
| 1143 | Qy | ACAGTTGATGAATTTGAAAAATATATCTCTGGTTCAAAAAGACGATTTGAGTTTTTAACCA | 1202 |
| 6285 | Db | ACAGTTGATGAATATGAAAAATGATAATTTCAACGACGAGCGCGCTTAGTTTATAGCCA | 6344 |

| | | | |
|----|------|----------------------------------------------------------------|------|
| Qy | 1203 | GGGATTACAGTCTCTCGCAGGTTAGTGGTCGTAGTAATATATACAGACTTCGACGCTA | 1262 |
| Db | 6345 | GGAACTACTGGTTTGGCAAAATATCTGGTAGAAAATATATATCTGATTTTGATGAATC | 6404 |
| Qy | 1263 | GTTCGGTTGGACTTTAGCATACATTAATAATGGACTATCTGGTCAGATATTAATAATTTTA | 1322 |
| Db | 6405 | GTAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTAAGATTTAT | 6464 |
| Qy | 1323 | TTAAAGCAGTGAAAGTTGTTATTTGAGAGAGGAAAGTAAGTAAAGTA----- | 1372 |
| Db | 6465 | CTCTAACACATAAAGGTAGTCTTACTTTGGGACAGTCTCTAAGTAAAGGTAAGTTTGAAA | 6524 |
| Qy | 1373 | -----TATGAAGTTTGTGGTCTCTTACTTTGGGACAGTCTCTAAGTAAAGGTAAGTTTGAAA | 1425 |
| Db | 6525 | GGATATAAATGAAATTTGTCTGGTTTCAAGTGGTGGTCTCTAGCACACCTTGAACC | 6584 |
| Qy | 1426 | TGTTAAACCGTTTGTGAAGGAAGAAAGCTTTTGGGTAAACATTTTGATAAAGAGATG | 1485 |
| Db | 6585 | TTTTGAAACCAATTTGGGAAATAAAGATAGGTTTGGGTAAACCTTTGATAAAGAGATG | 6644 |
| Qy | 1486 | CAAGAAGTCTTTTGAAGAATGAAAAATGTATCCATGTTACTTTTCCAAACAAATCGCAATC | 1545 |
| Db | 6645 | CTAGGATGTTTCTAAGAGAGAGATGTAATATCATTCCTTTTCCAAACAAACCGTAATG | 6704 |
| Qy | 1546 | TCATTAAATTTAGTGAAAAATCTTCTTAGCTTTCAAAAATTTTACGTGATGAGAAACGAG | 1605 |
| Db | 6705 | TCAAAACTTTGGTAAAAATACTATCTAGCTTTTAAAGTCTTTAGAAAAAGAAAGACGAG | 6764 |
| Qy | 1606 | ATGTTATTTATTTATCTGCTGGCGCTGCTGCTGCCCTCTTTTACATCGGAAACATAT | 1665 |
| Db | 6765 | ATGTTATCATATCATCTGGTGGCCCTGTAGCAGTACCATTCTTTTATTTGGTAAGTTAT | 6824 |
| Qy | 1666 | TTGAGCAAAAGACGATTTATATGAAGTATTTGATCGAGTTAATAATCTACATTAAGT | 1725 |
| Db | 6825 | TTGGTTGTAGACCGTTTATATAGAGGTTTTCGACGATAGATAAACCAACTTTGACAG | 6884 |
| Qy | 1726 | GAAACCTAGTTTATCCCGTAAACAGATATTTTATTTGTTAGTGGGAAAGAAATGAAGAG | 1785 |
| Db | 6885 | GAAATTTAGTGATCTCTGTAACAGATAAATTTATTTGTTAGTGGGAAAGAAATGAAGAG | 6944 |
| Qy | 1786 | TATATCTTAATCTATTAACTTGGGAGTATTTTAAATGATTTTCTGATAGTAAAGTGAAC | 1845 |
| Db | 6945 | TTTATCTTAGGCAATTTAATTTAGGAGAAATTTTAAATGATTTTCTGATAGTAAAGTGAAC | 7004 |
| Qy | 1846 | TCATGAACAAACAGTTTAAATCGATTGATAAAGAGATTTGATTTATTTGAAAAAATGAAG | 1905 |
| Db | 7005 | ACATGAACAGCAGTTCAACCGTCTTATTAAGAGTTGATAGATTAAGGGACAGTGC | 7064 |
| Qy | 1906 | TATACCGACGAAATATTTATTTCAACAGGATATTTCTGACTATATTTCCAGAAATTTGCAA | 1965 |
| Db | 7065 | TATTTGATCAAGAGTGTCTTCAACAGGGTTACTCAGACTTTTGAACCTCAGAAATTTGCA | 7124 |
| Qy | 1966 | GTATAAAAAATTTCTCAGTTTACAAGAAATGGAACATATATTAACAAATCAGAGTAGT | 2025 |
| Db | 7125 | GTGGTCAAAATTTCTCATATGATGATGAATCTTTATGAATGAAGAGTGAATGATGTT | 7184 |
| Qy | 2026 | TATTTGGCAGCGGCCCCGCTACTTTTATGAATTTCAATTTATCCAAAGGAAAAAACAAT | 2085 |
| Db | 7185 | TATCACACATGGCGGTCACGCGAGTTTATGAATGAGTCTTCAAGGGAAAAAACAAT | 7244 |
| Qy | 2086 | ATTGTTTCTAGACAAAAAAGTATGTTGAACATGTAATGATCATCATCAAGTAGGTTGT | 2145 |
| Db | 7245 | TGTGGTTCTTAGACAAGAACAGTTTGGAGAGCATGTGAATAATCATCAGGTGGATTTTT | 7304 |
| Qy | 2146 | AAGAAGAAATTTTACAAGATAATATTTTATTTATAGAAAAATATAGATGATTTGTTGA | 2205 |
| Db | 7305 | GAAAGAGTTATCTTGAATAATTTGAATTTAGATTTATTTTGAATATCATGTAATTTAGAA | 7364 |
| Qy | 2206 | AAAAATTTTGAAGTTTCTAAGCAAC-----TAACTTTTACATCAAAATTAATTTTTTTTG | 2262 |
| Db | 7365 | TATTTAAGGAAAAAATAATCTACTAGTTAAAGTAAATATATCAAAAAACAATGATTTTTG | 7424 |

| | | | |
|-----------------------------------------------------------------------------------------|------|----------------------------------------------------------------|------|
| Qy | 2263 | TGAAAGATTAAACAAATAGTTGAAAAATTTATAGGATCAAGAAAAATGAATAATAAA | 2322 |
| Db | 7425 | TTTCTCTTTCAAAAAATGAACATTTTCATAAACTATTTTGAATAAATATATTTTGTGGAGAA | 7484 |
| Qy | 2323 | AAGATGCATATTTTGATA | 2339 |
| Db | 7485 | AAAAATTTGAATTTAACA | 7501 |
| RESULT 9 | | | |
| US-10-461-990-4 | | | |
| : Sequence 4, Application US/10461990 | | | |
| : Publication No. US20040023361A1 | | | |
| : GENERAL INFORMATION: | | | |
| : APPLICANT: Nestec, S.A. | | | |
| : APPLICANT: Neeser, Jean-Richard | | | |
| : APPLICANT: Mollete, B. | | | |
| : APPLICANT: Stringeles, Francesca | | | |
| : APPLICANT: Zinc, Robert | | | |
| : APPLICANT: Kratky, Zoe | | | |
| : TITLE OF INVENTION: Lactic Acid Bacteria Producing Polysaccharide Similar to those in | | | |
| : FILE OF INVENTION: Human Milk and Corresponding Genes | | | |
| : FILE REFERENCE: 88265-10322 | | | |
| : CURRENT APPLICATION NUMBER: US/10/461,990 | | | |
| : CURRENT FILING DATE: 2003-06-16 | | | |
| : PRIOR APPLICATION NUMBER: US 09/548,606 | | | |
| : PRIOR FILING DATE: 2000-04-13 | | | |
| : PRIOR FILING DATE: 1998-10-09 | | | |
| : PRIOR APPLICATION NUMBER: CH 94 203245.2 | | | |
| : PRIOR FILING DATE: 2000-01-01 | | | |
| : NUMBER OF SEQ ID NOS: 37 | | | |
| : SOFTWARE: PatentIn version 3.2 | | | |
| : SEQ ID NO 4 | | | |
| : LENGTH: 18373 | | | |
| : TYPE: DNA | | | |
| : ORGANISM: Streptococcus macedonicus | | | |
| US-10-461-990-4 | | | |
| Query Match 9.8%; Score 684.8; DB 17; Length 18373; | | | |
| Best Local Similarity 49.0%; Pred. No. 1.5e-86; | | | |
| Matches 2838; Conservative 0; Mismatches 2767; Indels 183; Gaps 31; | | | |
| Qy | 498 | GATAAAATTAATTTATCATTTACCGCTCTATTTATCTGTGGAAGAGCTATAGAGTTTCA | 557 |
| Db | 6258 | GATAATCTCGATATTTTACAGCGCTATCTTAAAGAGTACTAAAGAAATTTAGGATTTTGTG | 6317 |
| Qy | 558 | ACAAGGGAGTGTGCGACACCGTCTTTATATAATCTACCAAGTCTAGTTTGTAGCGTAAAG | 617 |
| Db | 6318 | GCAAAATCAACAAGTTGACGAGATATTTGT---TGCCAAAAAATCACCAGATGATTTTAT | 6374 |
| Qy | 618 | CAATTCGTTTCAGATTTTTCAGTTTGTAGGTTTGTAGTATTTGATGATGATTAATTTCAATC | 677 |
| Db | 6375 | GAAATTTGCTCATTTGCTTAAATTTGTTAGGAAATTTCCACGACAGTAGCTGTTGGGAATAT | 6434 |
| Qy | 678 | GGTTTACTGCGTTGAAAAACAAAAATTCCACTGCTAGTGCACCATAGCATTTGTAAC | 737 |
| Db | 6435 | TCGGACTTCTATGTTGGAATAGTGTCTTAAAAAAAGTAGGTGATACGACCTTCATAACG | 6494 |
| Qy | 738 | TTTTCCACAAATTTTATAGCCCTAGTCATATCATGATGAACGACTTTTGGATATCTC | 797 |
| Db | 6495 | ACAGCATCAATTTGTAATTTCCGTCAGATTTGCTTTAAACCGTCTTATGGATATTTGCA | 6554 |
| Qy | 798 | GGAGCGGTAGTCGGGTAAATTTATTTGTTGGTATAGTTTCTATTTTGTAGTTTCAATTTAT | 857 |
| Db | 6555 | ATAGCTTTAGTTGGCTTAGTATTTACTGGTATTTGATGACCATTTATTCACACCGATATC | 6614 |
| Qy | 858 | CGTAGAGATGG---TGACCGGCTATTTTGTCTCAGAAACGAGTTTGGACAGATGACGC | 914 |
| Db | 6615 | AAGAAACAATCACCAGGACCTCTAAATCTTCAAAACAAAAACGTTGTGTTAAAAACGGTAAA | 6674 |
| Qy | 915 | ATATTTTACATCTTACAGTTTTCGATCGATGTTGTTGATGCTCAGGAGCGCAAAAAAGAC | 974 |

Db 6675 GTTTTGAATTTTACAAATTTTGAAGCATGTACACCGATGCCGAGAACGCCAATAAGAA 6734
Qy 975 TTGCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTTTAAATGGGAAAAACGATCTTA 1034
Db 6735 TTACTAACACAAAATGATTTGGATACCTGACTTAATGTTTAAAGTGG--ATGATGACCCCTC 6792
Qy 1035 GAAATTACTCCAATTGGCAATTTATACGCAAAAACAAAGTTTGAAGAGTTTACCAAGTTT 1094
Db 6793 GTATCTTCCCAATTTGGACATAAGTTACGTTGATGTTGCTCACTT-GATGAATTTACCACAAATTT 6851
Qy 1095 TATATGTTTAAATTTGGCGATATGATCTGTTGTCACAGTCCACCTTACAGTTGATGAA 1154
Db 6852 ATTAATGTCCTAAAGGTGAATGTCGTTGTTGGGCACAGTCCACCAACGCTTGACGAA 6911
Qy 1155 TTGAAAATATACCTCTGGTCAAAAGAGACGATTTGAGTTTAAACCAAGGATTTACAGGT 1214
Db 6912 TATCATCTATGATTTACATCACTTCAACCGATTCGACAAACCAACAGGAATTTACTGTT 6971
Qy 1215 CTCTGGCAGTTAGTGGTCTGATTAATATACAGACTTCGACGAGCTAGTTGGTGGAC 1274
Db 6972 TTATGGCAAGTTAGCGGTCTGATGACATTTACCGACTTTGAAGAGTCTGACCACTTGAT 7031
Qy 1275 TTAGCATATGTAATTTGACATCTGCTCAGATATTAATTTTAAAGACAGTG 1334
Db 7032 ATGAAGTATATCCAAAACCTGGAGCATCAGTGAAGATATTAATTTATTTGCCAAAACATTT 7091
Qy 1335 AAAGTTGATTTGTTGAGAGAGGAGTAAGTAA--AGTATATGAAGTTTGTGTTGGTCG 1392
Db 7092 GGAGTCGTACTAAAAGAGAGGGAAGTAAGTAGATATATTAAGAGTTTGTGTTAGTAG 7151
Qy 1393 GTTCTTCAGGGGACATTTGACTCACTTGTATTTGTTTAAACCCGTTTGGGAAGGAAG 1452
Db 7152 GTTCTTCTGTTGGACATTTGGCACATTTGAAATATGCTAAACCCCTTTTGGAGTGAACATA 7211
Qy 1453 AACGTTTTTGGGTAACTTTGATTAAGAGGATGCAAGAGCTTTTGAAGATGAATAA 1512
Db 7212 GCCGTTTCCGGGTACATTTGATTAAGAGAGCGCAAGAGTGTGTTAAGTCATGAATAA 7271
Qy 1513 TGATTCATGTTACTTTTCCAAACAAATCGCAATCTCATTAATTTAGTGAATAACTTTCT 1572
Db 7272 TTTATCCGTTATTTTCCGACTAACAGAAATTTAAGAAATTTGTAAGAACACTTTCT 7331
Qy 1573 TAGCTTTCAAAATTTTACGTGATGAGAAACAGATGTTTATTTATTTCACTCGTGGCGCG 1632
Db 7332 TAGCACTTGAATTTTAAAGAAAGAAACCTGACGTTTATTTATTTCACTAGCAGCAGCGG 7391
Qy 1633 TTGCTGCCCTCTTTTACATCGGAAACATTTTGGAGCAAGACGATTTATTTGAAG 1692
Db 7392 TAGCAGTTTCCATTTTATCTGGGTAACTGTTTGGAGCGAAACGGTTTATATCGAAG 7451
Qy 1693 TATTTGATCGAGTTAATAAATCTACATTAACCTGGAAACACTAGTTTATCCCGTACAGATA 1752
Db 7452 TATTTGATAGAAATAGATAAACCGACTGTGACTGGAAAGTTGGTTATCCAGTGACAGATA 7511
Qy 1753 TTTTATTTGTTCAGTGGGAAGAAATGAAGAAGGTATATCTTAAATCTATTAACTTTGGGGA 1812
Db 7512 AAATTTATTTGTTCAGTGGGAGGAGATGAACCTGCTATCCCAAGCTATTATCTGGGGA 7571
Qy 1813 GTATTTTAAATGATTTTGTAAACAGTAGAGAACTCATGAACAAACAGTTTAAATCGATGAT 1872
Db 7572 GTATTTTAAATGATTTTGTATACAGTTTGAACCTCATGAACAGCCCTTTTAAATAGGCTTAT 7631
Qy 1873 AAAAGAGATTCGATTTTAAAGAAATGGAAGATTAACCGACGAAATTTTATTTCAAC 1932
Db 7632 TAAGGAAGTTGATTCGTTTAAAGAAAGAGGATTTATTAACAGATGAGGTTTATTTACAG 7691
Qy 1933 AGGATTTCTGACTATATTTCCAGATATTTGCAAGTATATAAAATTTCTCAGTTACAAAGA 1992
Db 7692 AGGTTTTCAACTATATGAGCCTCATACTGTGACTGGAAATAATATTTCTTATCTGA 7751
Qy 1993 AATGAAACAATATATAAACAATACAGAGTAGTTATTTGCCAGGAGCCCGCTACTTTT 2052
Db 7752 AATGGAAGATTACATGAATCGTGCAGATATTTATTTATCACGCATGCTGGTCCAGCGACAT 7811

Qy 2053 TATGAATTCATTTTCCAAAGGAAAAAACAATTTATGTTTCTCTAGCAAAAAAAGTATGG 2112
Db 7812 CATGGGCAATTCGTTAAAGGAAAAAACCCGATGTTGTTTCCAGACAGGAAAAAGTTGG 7871
Qy 2113 TGAACATGTAATGATCATCAAGTAGAGTTTGTGAAGAGAATTT---TACAGATAATAA 2169
Db 7872 AGAGCATGTAATGATCATCAAGTTGATGTTGCTGACACAGGTTCTGACGATTTGGGAG 7931
Qy 2170 TATTTATTTATAGAAAATATAGATGATTTGTTTGAATAAATTTATGAGTTTCTTAAGCA 2229
Db 7932 TATCGTTGCTGAGAGAAATTAATGATTCGAAAATTTATTTTAAATTTAGATTTAATTTGT 7991
Qy 2230 AACTAACTTTATCATCAATAATAATTTTGTGAAAGATTTAAACAAAATAGTTGAAAA 2289
Db 7992 AGATGAAAGTTCCAAATTCGAAACCACTAAGATTTAATAGTCAATTTAAACCAAG---AAA 8047
Qy 2290 ATTTAATGAGATCAAGAAATGAATTAATAAAGATGCATATTTGATAATGGCTTATC 2349
Db 8048 TAGAAAGTTTGGTTAGATGAATGATTCCTAAAGATTCATTTATTTGGTTGGAGAA 8107
Qy 2350 ATAAATTTTCTCAGATTTTACTGGAGAGGATACAGATATTTATCATCTCTCTCAGGAGA 2409
Db 8108 ATCCTCTTCTGACAGTGTAAAAATTTGTATAAATTCGTGGAATAAATCTGTCCAAAT 8167
Qy 2410 ATGCACACCAATTTAGTTCCTTCAGAAATACCTGTATAATTTATTTTAAAT-ATTCTCAGGAT 2468
Db 8168 ATGAATAATCGAATGGAATGAATCAAAATTAATGATGATACATAAAATTCATATATTTCTG 8227
Qy 2469 TATATGTTGAATTTACAAAGATGAGCAAAAATATTAAGAAAAATAGGATATATGAAACGAG 2528
Db 8228 AAGCTTATAAATAAAGAAATATGCTTTGTTATCTGACTATGCTAGGCTAGATATCATAT 8287
Qy 2529 TTAATGTTTACAGATTTTCTCCTAATA-----TATCAGAAAAAACTATTGATAATGT 2580
Db 8288 ATATAGGGCGGGTTTTTATTTAGATCTGATGTTGAATTTGTTAAAGCATTTGGACGAT 8347
Qy 2581 ACTGTTTAGAATTTTATTAAGAAATGTATCGAGCTTTTGAATACATATTTTACAAAGATTTGT 2640
Db 8348 TAACTCTGAACTGTTATATATGGAATGGAACAAAGTGGTGGTCTGTTAACTCTGATTTAG 8407
Qy 2641 GTTTATGATAGATAAAAAACATGGTCTAAGAAATAGAAATTTGGTTCTAAATTTGGGTTTCG 2700
Db 8408 GTTTTGGTGCAGAA-AAAAGGACATCTTTTATAAAGAAAAATATGACGCAATATGAAGAA 8466
Qy 2701 CTTCACATGATTTTGTGGCAATCTTTTATCAAAATGAAGAAACGAAACAGCTTATTATTT 2760
Db 8467 GTTCTCTTAACTTTAAGCTACTAGAAACATGTGTGGATATCACGACAAATTTATTTATA 8526
Qy 2761 AAGTAATCTAAATGTCCAGATGAACATTTTATACAGACAAATTAAGAAAAATATGAAATTT 2820
Db 8527 TCAAAGGGTTATTTAGTAGAAAAATTCATATCAAAAAATTAGTGATGTCTCAATTTATCCA 8586
Qy 2821 TCAAATAGATTTATTAATATGGAATTTAAGATATATAAAGTGGAAAAAATCAACATCT 2880
Db 8587 ACAGATTTTTTTTGTCCGTTTAAATGATCAACACAAAGAAATGGGAATTAACATAAAATACT 8646
Qy 2881 TCTCCTATGCTTTACAGATGATTTCTATGATGAATTTGCTAAATGCAAGAAATTTAGGT 2940
Db 8647 TATTCAATTCATCATTTATGATTTCACTTGGTATG-----GTAATGGTGT 8691
Qy 2941 TTTTATTTGCTGAAAGTTTAAAAATGAAAAATTAATCTAAATTTTAAAGAAATTTATTTACT 3000
Db 8692 AGTCAATAATTAAGAAATTTATTTACCATTAAGATTTAAATCTCGTATCCTTATTCAT 8751
Qy 3001 AAAAAATAAATAGTGTGATTTTGTGAGAGTAAATGATGTTTAAATTTTAAATATGACC 3060
Db 8752 AAATATTTAGGTGAAGGCTCTTATGCTA-AAATCAAGCTATTTATTAAGAAATGATATTT 8810
Qy 3061 CGGAATATTTTATTTTAAAGTACTTCTGTTGATTTATTTTATTTCCAGAGCAAAAGTATG 3120
Db 8811 TTCAAAGAGGATATTTTGTAACTAAATTAATTTGATTTTGTGAT-----ATATATATATTT 8864

Db 10917 TTGATGATGGATCTACAGACAAATCAATATATCGATATGTAAGATGATGCTGCGATGAT 10976
Qy 5341 AAAGAGTAAAAATTTTTTTCATATCATAGTGGAGTATCAAAATGCTAGAAATCATGCAA 5400
Db 10977 CTCGAATTAAGCTTTTTCTAAAGAGATGGCGGTTTCTAGCGCTCGAATCTAGTGC 11036
Qy 5401 TAAAGCGAGTACAGCTGAATATATATGTTGTTGACTCTGATGATGTTGTTGATGATA 5460
Db 11037 TTCTACATGTTCAAGGAGAGTACGTTGTTGTTGATGATGATGATGATGATGATGATGAT 11096
Qy 5461 GATTAGTAGAAAAATTTATATTTTAAATATATAAAGTAGAAGTGAATTTATCTGTTGTT 5520
Db 11097 AATATTGTAACATTTATCACTTACTATTAAGTACTAAGTCAAGTGTAGCT----- 11149
Qy 5521 TGTAGCTACTTTTTTCAGAAAATATAAATAATTTTTGAAAGTGAATCAATCAATATTTGAT 5580
Db 11150 -TCTGTAAGTCGTTTATAACATTTTGAATAAAGAGTGTGAAGATATCGGATTTATCTT 11207
Qy 5581 TTGAAGCAATTAATACCGTGCAGGACATGGGAGAAAAAATTTTATGAAATTTGATATAA 5640
Db 11208 TTAATCAATAATCAATCAGATGAAGCTTTAAGAAAAATTTCTTTTAGGTGAGGGGATAAAT 11267
Qy 5641 ATAATATTTTTTCTACTCTCTGTTGTAACATATATAAGAAAGATACATACAGATCTTT 5700
Db 11268 GTTATCTTTTTTCAAAATATTTAAATATGAACATATAAAGGACTCCGATTTGATGAAA 11327
Qy 5701 TTCAAGAAATCAATGGTTAGGAGAAATTTACTTTTTTAATCTGCATTTATTAAGAAATA 5760
Db 11328 GTTTAGATCAGCAGAGACGTTTTGTTTATTTATCAACCTTTAAGAACATAAATTTTG 11387
Qy 5761 TAGATAGAGTTAGTTATTTGACTGAACATCTTTATTTTTATAGGAGAGGTATCTAAGTA 5820
Db 11388 CATCTATGATGGCACTGTCAGATTTATTTTATATCTTTA-GAAGAGGATCTTTAACA 11446
Qy 5821 CAGTAAATCTTTTAAAGAGGTGTGTTTTGCAATTTGGAAATTTGCAAAAACAGTGA 5880
Db 11447 AATAAAGACTGACTTCATCAAGAAATGATATGATGTTCCATTTAGGTTGCGGAATTTTACT 11506
Qy 5881 TAGTATCTTTTAAAGCAATATATGCTGAGGATTTGACGTATCAATTTGTTAAAGATATA 5940
Db 11507 AGAGTTGCAACGCAACAAAAATGAAATGTTAAGTGA--AATTAATGAATATCAT 11564
Qy 5941 TACGTGCGCAAGTATTTTATTTATAGCTTACTAATGTTTAAATACGGAACACAGTCTATTT 6000
Db 11565 TAAAGGTGAGGTTCTTCTGAGTGGATTTCAATTAATAGTGAACCTTAGAATTCGATTTGAAG 11624
Qy 6001 TTGCAAAATTTTAAATTTTGAATCTTTTAAATAATATATTTTAACTTTGTTAAAG 6060
Db 11625 AATATTATTAATCATACTGAGAGA-AGTTAGAAGTTTAAATTTGTTACATAAAGTTTCAA 11683
Qy 6061 TATCTCAAAAAATTTCTTTGTTCTAAAAATTTTGTATAGAATTTGTTTGAACA----- 6114
Db 11684 TATCTAACTTTTAAAAAATTTATAGGATTTATTTTAAAAAGTTAGTCTTAGATGTT 11743
Qy 6115 AAGTTTTTAAAAAATTTATGTTTATATAGGAAGATATCATGATACTATTTAGTAA-- 6172
Db 11744 ACAATCTTAAAAAATAAATAGTATCTCGTGAAGAGTATTCATGATTTTAAATAGTAACC 11803
Qy 6173 ----AATTTCTAATTTGACTATATATATATAGTAAAAATATTTATCTAAAATGATAG 6228
Db 11804 CTCTGTTTCAATTTATTTCAATTTTATAATGTAAGAAATTTTAAAGACAGTGCTCTA 11863
Qy 6229 ATAGATTTGAATCAGACCTACAACA 6256
Db 11864 CTTGAGTGTTCACATAATTTATGACA 11891

RESULT 10

US-10-804-408-166

; Sequence 166, Application US/10804408

; Publication No. US20040253617A1

; GENERAL INFORMATION:

; APPLICANT: PANRONG, KONG

; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCI/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 166
; LENGTH: 2226
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-166

Query Match 9.3%; Score 652.4; DB 18; Length 2226;

Best Local Similarity 59.6%; Pred. No. 2.6e-82;

Matches 1225; Conservative 0; Mismatches 791; Indels 38; Gaps 6;

Qy 16 TGGCAATTTATGATGATAGTGGTGGCAATTTCTGCAATCTTAAACAAGTCATATACCAA 75
Db 182 TAGGATGATTTCAAAACAGTTGTGTTTATTTTCTGCAAGTTTGACATTAACATTAATA 241
Qy 76 ATGCTGATTTAAATCG---TTCTGGAAATTTTATCATAATGATGTTTCATTTTTCAT 132
Db 242 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATTTATTTGCT 301
Qy 133 TTTTATATCTCGTATGCGAGTTGAAATTTGATGATAGAGTAACTGATGAGTTTGAATA 192
Db 302 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCTGGCTATCTTGAAGAGTTTAAAA 361
Qy 193 AAACATTTAACTATAGTATATATTTGCAATTTTCTTACGCGAGTATCATTTTGTGG 252
Db 362 TGGTATTGAATACAGCTTTTACTATATTTTCAATCAAGTTTCATTTTATTTTATTTTA 421
Qy 253 AGAATAATTTTCGCACTTTTCAAGAGTGGTGGCGGTATTTTCCACATTAATAAACTTCGTT 312
Db 422 AAACTCATTTTACATGACAGACTTCTCTTTTCTTTTATTTGCTATGATTCGATTT 481
Qy 313 TGGTATACCTTATTTAAAGTAAATTAATTAAGCAGTTTAAGGATAGCTTTCTATTTTCGACA 372
Db 482 TATTATATCTATTGAAATTCATTTTAAATAATTTTAAATAATTTCTTACGCTAAGTTT 541
Qy 373 TCTATCAAAAAAGACGATTTCTAATTAACGCGTACAGATGGGAATATGCAAGTTT 432
Db 542 CACGAGATACCAAGTTGTTTGTATAACGAATAAGGATTTCTTATCAAAAATGACCTTTA 601
Qy 433 TATTTGAATCACATAAAACAATTTCAAAAAATCTTGTGCAATGTTAGTATTTTAGGTACAG 492
Db 602 AGAATAAT---ACGACCAATTAATATCGCTGCTGTATCTTGGACTCTCTGAAAAGG 658
Qy 493 AAATAGATAAATTAATTTTATCATTAACGCTCTATTTTCTGTGGAAGAAGCATATAGAT 552
Db 659 ATTGTTATGATTTGAAAACATAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 718
Qy 553 TTTCAACAAAGGAAGTGTGACCAAGCTCTTTTAAATCTTACCAAGTGAAGTTTATAG--- 609
Db 719 AGTTAACTGCTTAACTGTTGATCAAGCTTTTATTAACATACCCATTTGAAATTTATTTGTA 778
Qy 610 -----ACGTTAAAGCAATTTGTTTTCAGATTTTGTAGTTTGTGATGTTAGCGGTG 663
Db 779 AATACCAATTAACAGATTTATTAATGACATTAAGCAATGAGGATGATTTGTCATGTTA 838
Qy 664 ATATTAAATTCATCGGTTTCTGCGTTGAAAAAACAATAAATCAACTGCTAGGTGACC 723
Db 839 ATGTAGAGGCACTTAGCTTTTGATAATATAGGAAAAAGCGAATCCAAACTTTTGAAGAT 898
Qy 724 ATAGCATTTAACTTTTTCACAAATTTTATTAAGCCTAGCTCATATCATGATGAACGAC 783
Db 899 ATAGTGTATTACATAATTTCTATGAATTTCTATAAATATAGTACCTTTATAGCAAAACGAT 958

Db 602 GGAATAAT--ACGACCAATAATATATCGCTGTCTGTATCTTGACCTCTCTGAAAAGG 658
Qy 493 AAATAGATAAAATAATTAATTAATACCGCTCTATTAATCTGTGGAAGAAGCTATAGAGT 552
Db 659 ATTGTTATGATTTGAAAACATAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 718
Qy 553 TTTCAACAAGGAAGTGTGCGACGAGCTTTTATAACTACCAAGTGAAGTTTGTAG--- 609
Db 719 AGTTAACTGCTTAATCTGTTGATCAAGCTTTTATTAACATACCCATTGAATTAATTTGGTA 778
Qy 610 -----ACGTAAAGCAATTCGTTTCAGATTTTGAAGTTTGTAGTATTTGAATGAAGCGTTG 663
Db 779 AATACCAATACAGATTAATTAATGACATTTGAAGCAATGGAAGTGAATGTCATGTTA 838
Qy 664 ATATTAATTCATTCGGTTTACTCGGTTGAAAAAACAATAATCCAACCTGCTAGGTGACC 723
Db 839 ATGTAGAGCACTTAGCTTTGATAATATAGGAGAAAGCGAATCCAAACTTTTGAAGGAT 898
Qy 724 ATAGCAATTTGTAATTTTCCAAATTTTATAGCCTAGTCAATATCATGATGAACGAC 783
Db 899 ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT 958
Qy 784 TTTTCGATATACTCGGAGCGGTAGTCGGTTAAATTTATTTGTTGATAGTTTCTATTTTGT 843
Db 959 TTTTGGATATACCGGTGCTATATAGGTTTCTCATATGTGGCAATGTGGCAATTTTTC 1018
Qy 844 TAGTTCCAAATTTATCGTAGATGTGGACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC 903
Db 1019 TAGTTCCAAATCAGAAAAGATGTGGACCGGCTATCTTTTCTCAAAATAGAGTAGTTC 1078
Qy 904 AGAATGGAGGCAATTTTACATCTCAAGTTTCGATCGATGATGTTGATGCTGAGGAGC 963
Db 1079 GTAATGTTAGGATTTTATAGTTCTATAAATTCAGATCAATGCGAGTAGATGCAAGCAAA 1138
Qy 964 GCAAAAAGACTGTCTGACCAAAACAGATGCAAGGTTGGTATGTTTAAATGGGAA 1023
Db 1139 TTAAGAAGATTTATAGTTCAATCAATCAATGACAGGG-----CTAATGTTTAAAGTTAGA 1193
Qy 1024 AAACGATCCTAGAAATTAATCCTAAATTTGGACATTTCTATAGCGAAAAACAAGTTTAGACGAGT 1083
Db 1194 AGATGATCCTAGAAATTAATTAATAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT 1252
Qy 1084 TACCAAGTTTATTAATGTTTAAATTTGGGATATAGAGTCTAGTTGGTACAGTCCACCTA 1143
Db 1253 TGCTCAATTTCTATAATGTTTAAAGGTGATATGAGTTTAGTAGGACACGCGCTCCCA 1312
Qy 1144 CAGTTGATCAATTTGAAAAATATACCTCTGCTCAAAAGAGAGAGTGTAGTTTAAACGAG 1203
Db 1313 CAGTTGATGAATATGAAAAGTATAATTCACGCGAAGCGAGCGCTTAGTTTAAAGCCAG 1372
Qy 1204 GGATTAACAGGTTCTGCGCAGGTTAGTGTGCTAGTAATATACAGACTTTCGACGACGTAG 1263
Db 1373 GAATCAGCTGGTTTGGGCAATATCTGGTAGAATAATATCACTGATTTGATGAATCG 1432
Qy 1264 TTCGGTTGGACTTAGCATACTGATAATTTGGACTATCTGCTCAGATATTAATAATTTTAT 1323
Db 1433 TAAAGTTAGATGTTCAATATATCAATGAATGCTCTATTTTGGTCAGATTAAGATTTTC 1492
Qy 1324 TAAAGACAGTGAAGTTGTTGTTGAGAGGGAAGTAAAGTAT----- 1373
Db 1493 TCCTAACACTAAAGGTAGTCTTACTTGGGACAGGTGCTAAGTTAAAGGTTAAGGTTTGAAG 1552
Qy 1374 -----ATGAAAGTTTGTGTCGTTCTTCAGGGGACATTTGACTCACTCTATTT 1426
Db 1553 GAATATAATGAAATTTGCTGTTGTTTCAAGTGGTGCATCTAGCACACTTGAACCT 1612
Qy 1427 GTTAAACCGTTTTCGAAAGGAAGAAACGTTTTCGGGTAAATTTGATAAGAGGATGC 1486
Db 1613 TTTGAAACCCATTTGGGAAAAGAGATAGGTTTTCGGGTAACTTTGATAAAGAGATGC 1672
Qy 1487 AAGAGTCTTTGAAAGATGAAAAATGATTCATGTTTACTTTTCAACAAATTCGCAATCT 1546
Db 1673 TAGGAGTATCTTAAGAGAAGAGATTGTATATCATTTGCTTCTTTCACAAACCGGTAATGT 1732

Qy 1547 CATTAAATTTAGTGAAAAATACCTTTCTTAGCTTTCAAAATTTTACGTGATGAGAAACCAGA 1606
Db 1733 CAAAAACTTGGTAAAAAATACCTATTCTTAGCTTTTAAAGTCTCTTAGAAAGAAAGACCAGA 1792
Qy 1607 TGTTATTAATTTCACTGCTGGTGGCGGCTGCTGCTCCCTTCTTTTACATCGGAAACTATT 1666
Db 1793 TGTTATCATATCATCTGCTGGCGCTGTAGCAGTACCAATCTCTTTTATATTTGGTAAGTTAT 1852
Qy 1667 TGGACAAAGACGATTTATATTTGAAGTATTTGATCGAGTTAATAATCTACATTAACCTGG 1726
Db 1853 TGGTTGTAAGACCGTTTATATAGAGTTTTCGACAGGATAGATAAACCAACTTTGACAGG 1912
Qy 1727 AAACTAGTTTATCCCGTAAACAGATATTTTATTTTTCAGTGGGAGAAATGAAGAAGGT 1786
Db 1913 AAAATTAGTGTATCTGTAAACAGATAAATTTATTTTCAAGTGGGAGAAATGAAAAAGT 1972
Qy 1787 ATATCTTAATCTATTAACTTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGAACT 1846
Db 1973 TTATCTTAAGCAATTAATTTTAGGAGGAATTTTAAATGATTTTGTCAAGTGGGACA 2032
Qy 1847 CATGAACAAACAGTTTAAATCGATTGATAAAAGAGATTGATTTTATCGAAAAAATCGAAGT 1906
Db 2033 CATGAACAGCAGTTCAACCGTCTTATTAAGAAGTTGATTAAGAAGGACAGGTGCT 2092
Qy 1907 ATAAACCGACGAAATATTTTATTAACACAGGATATTTCTGACTATATTTCCAGAAATATTGCAAG 1966
Db 2093 ATTGATCAAGAAGTGTTCATTTCAACCGGTTACTCAGACTTTGAACCTCAGAAATTGTCAG 2152
Qy 1967 TATAAAAAATTTCTCAGTTTACAAAGAAATGGAACAATATATTAACAATTCAGAAAGTAGTT 2026
Db 2153 TGGTCAAAATTTCTCATATGATGATGAATGAATCTTACATGAAGAAGCTGAGATTGTT 2212
Qy 2027 ATTTGCCACGAGG 2040
Db 2213 ATCACATGCGG 2226

RESULT 12

US-10-804-408-162
; Sequence 162, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; PRIOR FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 162
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-162

Query Match 9.3%; Score 650.8; DB 18; Length 2217;
Best Local Similarity 59.6%; Pred. No. 4.3e-82;
Matches 1224; Conservative 0; Mismatches 792; Indels 38; Gaps 6;
Qy 16 TGGCATTTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAACAAAGTCATATACCAA 75
Db 173 TAGCGATGATTTCAACGTTGCTAGTTTATTTTCTGCAAGTTTGACATTAACATTAATTA 232
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCATTAATGATGTTTCATTTATTTTGCAT 132
Db 233 CTCCTCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTTATGATACATTAATATTTGTT 292

;; PRIORITY APPLICATION NUMBER: AU PR 7749
;; PRIOR FILING DATE: 2001-09-19
;; NUMBER OF SEQ ID NOS: 182
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 173
;; LENGTH: 2226
;; TYPE: DNA

;; ORGANISM: Artificial Sequence

;; FEATURE:

;; OTHER INFORMATION: Consensus sequence

US-10-804-408-173

Query Match 9.3%; Score 649.2; DB 18; Length 2226;

Best Local Similarity 59.5%; Pred. No. 7.3e-82;

Matches 1223; Conservative 0; Mismatches 793; Indels 38; Gaps 6;

| | | | |
|----|-----|----------------------------------------------------------------|------|
| Qy | 16 | TGGCAATATTGATGATAGCAGTTGCAATTTCTGCAATCTTAACAGTCAATATACCAA | 75 |
| Db | 182 | TAGCGATGATTCACACAGTTGTTGTTATTTTCTGCAAGTTTGACATTAACATTAATTA | 241 |
| Qy | 76 | ATGCTGATTTAAATCG---TTCTGGAAATTTTATCATATGATGGTTCAATTTTGTGAT | 132 |
| Db | 242 | CTCCCAATTTTAAAGCAATAAGATTTATTTGTTGTTCTATTGATACATTTATTTGCT | 301 |
| Qy | 133 | TTTTTATCTCGTATGCGAGTTGAATTTGATGATAGAGTAACTGATAGATTTGAAA | 192 |
| Db | 302 | TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA | 361 |
| Qy | 193 | AAACATTTAACTATAGTATATATTTGCAATTTTCTTACGCGAGTATCATTTTGTGG | 252 |
| Db | 362 | TGGTATTGAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTTTATA | 421 |
| Qy | 253 | AGATTAATTTTCCGACTTTTCAAGAGCTGGTGGCGGTATTTTACATTAATAAATCTCGTTT | 312 |
| Db | 422 | AAACTCTTTTACACGACAGCTTCTCTTTTACTTTTATTTTATTTGCTATGATTCGATTT | 481 |
| Qy | 313 | TGGTATACCTATTAACGTAATATTAAGCAGTTTAAGGATGCTTCTATTTTCGACAA | 372 |
| Db | 482 | TATTATATCTATGAAATTCATTTTAAATATTTATCGAATAATTTCTTACGCTAAGTTT | 541 |
| Qy | 373 | TCATCAAAAAAGACGATTTCTAATTACACGGCTGAACGATGGGAATATGCAAGTTT | 432 |
| Db | 542 | CACGAGATACCAAGTTGTTTGTGATACGAATAAGGATTTCTTATCAAAAAATGACCTTTA | 601 |
| Qy | 433 | TATTTGAATCACATAACAAATTCAAAAAATCTTGTTCATTTGTTAGTATTTTGTATACAG | 492 |
| Db | 602 | GGATTAAT---ACGACCTAATATATATCGCTGCTGATCTTGTGACCTCTGAAAAGG | 658 |
| Qy | 493 | AAATAGATAAAATTAATTTATCATTTACCGCTCTATTTATTTCTGTGGAAGAGCTATAGAGT | 552 |
| Db | 659 | ATTGTTATGATTTGAAACATAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG | 718 |
| Qy | 553 | TTTCAACAGGAGTGTGACACGCTCTTTTATAAATCTACCAAGTGAAGTTTGTAG--- | 609 |
| Db | 719 | AGTTAACCTGCTTAACCTGTGATCAAGCTTTTATTAACATACCCCATTTGAATTTTGGTA | 778 |
| Qy | 610 | -----ACGTAAGCAATTTCTGATTTTGTAGTTTGTAGTTTGTAGTTTGTAGTGAAGCGTTG | 663 |
| Db | 779 | AATACCAATACAGATATTAATGACATTTGAAGCAATGGAGTGAATTTGCTCAATGTTA | 838 |
| Qy | 664 | ATATTAAATTCATTCGGTTTACTGCGTTTGAATAAACAATAAATCCAACTGCTAGTGACC | 723 |
| Db | 839 | ATGATAGGCACTTAGCTTTGATATATAGGGAAGGGAATCCAACTTTTGAAGAT | 898 |
| Qy | 724 | ATAGCATTTGTAACCTTTTCCAAATTTTATAAGCCCTAGTCAATATCATATGATGAACGAC | 783 |
| Db | 899 | ATAGTGTATTACATATTCTATGAAATTTCTATAAATATAGTCACTTATAGCAAAACGAT | 958 |
| Qy | 784 | TTTTTGGATATCTCGGAGCGGTAGTCGGTTAAATTTTGTGGTATAGTTTCTATTTTGT | 843 |
| Db | 959 | TTTTTGGATATCACGGGTGCTATATAGGTTTGTCTCATATGTGGCAATTTTTC | 1018 |
| Qy | 844 | TAGTTCCAAATTTATTCGTAGATGTTGGACCGGCTATTTTGTCTCAGAAAACGAGTTGGAC | 903 |

| | | | |
|----|------|--------------------------------------------------------------------|------|
| Db | 1019 | TAGTTCCGCAAAATCAGAAAAGATGGTGACCGGCTATCTTTTCTCAAAATAGAGTAGGTC | 1078 |
| Qy | 904 | AGATGAGGCGCATATTTACATTTCTACAAAGTTTTCGATCGATGTATGTTGATCTCAGGAGC | 963 |
| Db | 1079 | GTAAATGTTAGGATTTTGTAGATTTCTTAAATTCAGATCAATCGGAGTAGATCGAGAACAA | 1138 |
| Qy | 964 | GCAAAAAAGACTTCTCTCAGCCAAAACAGATGCAAGGGTGGGTATGTTTTAAAATGGGAA | 1023 |
| Db | 1139 | TTAAGAAAGATTTTATAGTTTCACAATCAATGACAGGG-----CTAATGTTTAAAGTTAGA | 1193 |
| Qy | 1024 | AAACGATCTAGAAATTAATCTCAATTTGGACATTTCTATACGCAAAAACAGTTTTCAGCAGT | 1083 |
| Db | 1194 | CGATGATCTAGAAATTAATTAATAAGGAAATTTATTCG-AAAAACAAGCATAGATAGT | 1252 |
| Qy | 1084 | TACCACAGTTTATAATGTTTAAATTTGCGGATATGAGTCTAGTTGGTACACGTCACACTA | 1143 |
| Db | 1253 | TGCTCTCAATTTCTAATGTTTAAAGGTGATATGAGTTTAGTAGGACACACGCTCCCA | 1312 |
| Qy | 1144 | CAGTTGATGAATTTGAAAAATATATCTCTGGTCAAAAGAGAGAGATGAGTTTAAACGAG | 1203 |
| Db | 1313 | CAGTTGATGAATATGAAAAAGTATAATTCACCGCAGAGCGAGCTTAGTTTAAAGCCAG | 1372 |
| Qy | 1204 | GGATTACAGGCTCTGCGCAGGTTAGTGGTCTGATGATATATCAAGACTTCCAGCAGCTAG | 1263 |
| Db | 1373 | GAATCACTGGTTTGTGGCAATATCTGGTAGAAATATATTAATTAATTAATTAATTAATTA | 1432 |
| Qy | 1264 | TTCTGTTTGGACTTAGCATATCAATTCGATATTTGGAGCTATCTGGTCAAGATTAATAATTTTAT | 1323 |
| Db | 1433 | TAAAGTTAGATGTTCAATATATCAATGATGCTATTTTGGTCAAGATTAATTAATTAATTA | 1492 |
| Qy | 1324 | TAAAGACAGTGAAGTTGTTTGTGAGAGAGGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT | 1373 |
| Db | 1493 | TCCTAACACTAAAGGTAGTCTTACTTGGGACAGGAGCTAAGTAAAGGTAAAGGTAAAGGT | 1552 |
| Qy | 1374 | -----ATGAAAGTTTGTGGTGGTCTTCTGAGGGGACATTTGACTCAGTTGATTTT | 1426 |
| Db | 1553 | GAATATAATGAAATTTTGTCTGGTTGGTTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT | 1612 |
| Qy | 1427 | GTTAAAAACCGTTTGGGAGGAAAGAAACGTTTGGTGGTAAACATTTGATAAAGAGGATGC | 1486 |
| Db | 1613 | TTTGAACCCCATTTGGGAAAGAAAGATAGAGTTTGGTAAACCTTTGATAAAGAGATGC | 1672 |
| Qy | 1487 | AAGAGGCTTTTGAAGAAATGAAAAAATGATTCATGTTTACTTCCATTTCCAAACAAATCGCAATCT | 1546 |
| Db | 1673 | TAGGAGTATCTTAGAGAGAGAGATTTGATATATCATTTGCTTCTTCCACAAACCGTAAATGT | 1732 |
| Qy | 1547 | CATTAAATTTAGTGAATAATCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAACCCAGA | 1606 |
| Db | 1733 | CAAAAACTTGGTAAAAAATACTATTTCTAGCTTTTAAAGTCTCTAGAAAGAAAGAACCCAGA | 1792 |
| Qy | 1607 | TGTTTATTTATCTCTGGTGGCGGCTTGTGCTCCCTTCTTTTACATTCGGAACACTATT | 1666 |
| Db | 1793 | TGTTATCATATCATCTGGTGGCGCTGTAGCAGTACCATTTCTTTTATATTTGGTAGGTTAT | 1852 |
| Qy | 1667 | TGAGACAAAGACGATTTATATTTGAAGTATTTGATCGAGTTAATAATCTACATTAACCTGG | 1726 |
| Db | 1853 | TGGTTGTAAGACCGTTTATATAGAGTTTTCGACAGGATAGATATAACCACTTTGACAGG | 1912 |
| Qy | 1727 | AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTTCAGTGGGAGAAATGAGAAAGGT | 1786 |
| Db | 1913 | AAAAATAGTGTATCTCTGTAACAGATAAATTTATTTGTTTTCAGTGGGAGAAATGAAAAAAGT | 1972 |
| Qy | 1787 | ATATCTTAATCTATTAATCTTGGGAGTATTTTGTAAATGATTTTGTAAACAGTAGGAACT | 1846 |
| Db | 1973 | TTATCTTAAGCAATTAATTTAGGAGGAAATTTTAAATGATTTTGTCTCAGAGTGGGGACA | 2032 |
| Qy | 1847 | CATCAACAAACAGTTTAAATCGATTCGATTAAGAGAGATTTGATTTTATTTGAAAAAATGCGAAGT | 1906 |
| Db | 2033 | CATGACACGAGTTTCAACCGCTTTTAAAGAGAGTTGATAGATTTAAAGGACGAGTCTCT | 2092 |
| Qy | 1907 | ATAACCGACGAAATTTATTTTCAACAGGATATTTCTGACTATATTTCCAGATAATTTGCAAG | 1966 |

Db 2093 ATTGATCAAGAGTGTTTCATTCAAACGGGTCTACTCAGACTTTGAAACCTCAGAAATTTGTCAG 2152
Qy 1967 TATATAAAATTTCTCAGTTACAAAGAAATGGAACAATATATATTAACAATCAGAATGATTT 2026
Db 2153 TGGTCAAAATTTCTCATATGATGATGAATGAATCTTTACATGAAGAAGCTGAGATTTGTT 2212
Qy 2027 ATTTGCCACGGAGG 2040
Db 2213 ATCACATGGCGG 2226

RESULT 14
US-10-804-408-171
; Sequence 171, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 171
; LENGTH: 2226
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
; US-10-804-408-171

Query Match 9.3%; Score 648.8; DB 18; Length 2226;
Best Local Similarity 59.5%; Pred. No. 8.3e-82;
Matches 1222; Conservative 1; Mismatches 793; Indels 38; Gaps 6;

Qy 16 TGGCATTATTTGATGATAGCAGTTGCAATTTCTGCAATCTTTAAACAAGTCATATACCAA 75
Db 182 TAGCGATGATTCAAACMGTTGTGGTTATTTCTGCAAGTTTGACATTAACATTAATTA 241
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATCAATGATGTTTCATATTTTGCAT 132
Db 242 CTCCTCAATTTTAAAGCAATAAAGATTTATGTTGTTCTTATGATACATATATTTGCT 301
Qy 133 TTTTATATCTCGTATGCCAGTTGCAATTTGAGTATAGAGTAATCTGATAGAGTTTGAAA 192
Db 302 TTATCTTTCTGATTTTACAGACCTTTTGGAGTCGTGGCTATCTTGAAGAGTTTAAAA 361
Qy 193 AAACATTTAACATATAGTATATATTTTGCATTTTCTTACGCGAGTATCAATTTTGTGG 252
Db 362 TGGTATTTGAAATACAGCTTTTACTATATTTTCAATCAAGTTTCATTTTATTTTATTTTA 421
Qy 253 AGAATAATTTCCGACTTTCAAGCGTGGTCCGTGATTTTACATTAATTAATCAATCTCGTT 312
Db 422 AAAAATCTTTTACACGACGACTTTTCCTTTTACTTTTATTTGCTATGATGAATTCGATTT 481
Qy 313 TGGTATACCTATTTTAAACGTAATTAAGCAGTTTAAAGGATAGCTTTCTATTTTCGACAA 372
Db 482 TATTATATCTATGAAATTTCAATTTTAAATATTAACGAATATTCGAATATCTTACGCTAAGTTT 541
Qy 373 TCTATCAAAAAGACGATTTCAATTTCAACGGCTGAACGATGGGAAATATGCAAGTTT 432
Db 542 CACGAGATACCAAGTTGTTTGTATACGAATAAGGATCTTTATCAAAATGACCTTTA 601
Qy 433 TATTGATCACATAAACAATTTCAAAAATCTTTGTCATTTGGTATTTAGGTACAG 492
Db 602 GGAATAAAAT---ACGACCAATATATATCGCTGCTGATCTGATTTCTCTGAAAAGG 658
Qy 493 AAATAGATAAAATTTAATTTATCAATTTACCGCTCTATTTTCTGTCGGAAGAGCTATAGT 552
Db 659 ATTGTTATGATTTGAAACATCACTCGTTAAGGATAATAACAAAGATGCTCTTACTTCAG 718

Qy 553 TTTCAACAGGAGTGGTCCAGCAGTCTTTATATAATCTACCAAGTGAGTTTTTAG--- 609
Db 719 AGTTAACTTCGCTTAACCTGTTGATCAAGCTTTTATTAACATACCCATGGAATTAATTTGGTA 778
Qy 610 -----ACGTAAAGCAATTCGTTTCAGATTTTGAGTTGTTAGGTAATGATGAAGCGTTG 663
Db 779 AATACCAATATACAGATATTTAATGACATTTGAGCAATGGGAGTATGTCATGTTA 838
Qy 664 ATATTAAATTCATTGCGGTTTTTACTGCGTTGAAAAACAAAAAATCCAACTGCTAGGTGACC 723
Db 839 ATGTAGAGGCACCTTAGCTTTGATAATATAGGAGAAAAAGCGAATCCAAAACCTTTTGAAGGAT 898
Qy 724 ATAGCATTTGTAACCTTTTCCACAAATTTTATAAGCCTAGTCAATCATCATGATGAACGAC 793
Db 899 ATAGTGTATTATACATATCTTATGAATTTCTATAAATATATAGTCACTTTATAGCAAAACGAT 958
Qy 784 TTTTGGATATACTCGGAGCGGTAGTCGGGTAAATTTTGTGTTATGTTGTTTCTATTTTTGT 843
Db 959 TTTTGGATATACGGGTGCTATTATAGGTTTGTCTATATGTGGCATTTGGCAATTTTTTC 1018
Qy 844 TAGTTCCAAATTTCTGAGAGATGGTGGACCGGCTATTTTGTCTAGAAACGAGTTGGAC 903
Db 1019 TAGTTCCACAAATCAGAAAAAGATGGTGGACCGGCTATCTTTCTCAAAATAGAGTAGGTGTC 1078
Qy 904 AGAATGGACGCATATTTTACATTTCTCAAGTTTTCGATCGATGATGTTGATGCTGAGGAGC 963
Db 1079 GTAATGGTAGGATTTTGTAGTTCTATATAAATTCAGATCAATGCGAGTAGATGCAAAACAA 1138
Qy 964 GCAAAAAGACCTTGTCTCAGCCAAACCCAGATGCAAGGTTGGGTATGTTTAAAAATGGAA 1023
Db 1139 TTAAGAAAGATTTATTTAGTTTCAATCAATGACAGGG-----CTAATGTTTAAAGTTAGA 1193
Qy 1024 AAACGATCCTAGAAATTTACTCCAATTTGGACATTTTATACGCAAAAAACAAAGTTTACAGAGT 1083
Db 1194 CGATGATCCTAGAAATTTACTAAAATAGAAAAATTTATTTCG-AAAAACAAGCATAGATGAGT 1252
Qy 1084 TACCAAGTTTATATAATGTTTTTAATTTGGCGATATGAGTCTAGTTGGTACACGTCACCTTA 1143
Db 1253 TGCCTCAATTTCTATATATGTTTTTAAAGGTTGATATGATGTTTGTAGGAAACACGCGCTCCCA 1312
Qy 1144 CAGTTGATCAATTTGAAAAATATCTCCTGGTCAAAGAGACGATTTGAGTTTAAAAACGAG 1203
Db 1313 CAGTTGATGAATATGAAAAAGATTAATCAACGCAAGAGCGAGCCCTTAGTTTAAAGCCAG 1372
Qy 1204 GGAATACAGGCTCTCGCGCAGGTTAGTGGTCTAGTAAATATACAGACTTTCCAGACGCTAG 1263
Db 1373 GAATCACTGGTTTGGCAAAATATCTGGTAGAAATATATTAATCTGATTTTGTAGAAATCG 1432
Qy 1264 TTCGTTGGACTTAGCATACATTTGATAATTTGGAATATCTGGTCAGATATTTAAAAATTTTAT 1323
Db 1433 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTTGGTCAGATATTTAAGATTTATTC 1492
Qy 1324 TAAAGACAGTGAAGTTGTTATTTCTCAGAGAGGGAAGTAGTAAAGATAT------ 1373
Db 1493 TCCTAACACTAAAGGTAGTTTACTCGGACAGAGGCTAAGTAAAGGTAGGTTTGAAG 1552
Qy 1374 -----ATGAAAGTTTGTGTTGGTCCGTTCTTCAGGGGACATTTTGACTCAGTTTATTTT 1426
Db 1553 GAATATAATGAATAATTTTGTGTTGTTCAAGTGGTGGTCACTAGCACACTTTGAACCT 1612
Qy 1427 GTTAAACCGCTTTTGAAGAGAGAAAGCGTTTTTGGGTAAACATTTGTATGAAGAGGATGC 1486
Db 1613 TTTGAAACCCATTTGGGAAAAAGAGATAGGTTTTTGGGTAAACCTTTTGTATGAAGAGATGC 1672
Qy 1487 AGCAAGCTTTTGAAGAAATGAAAAATGATCCATGTTAATTTTCCCAACAAATCGAATCT 1546
Db 1673 TAGGAGTATTTCTAAGAGAGAGATTTGATATCAATGCTCTTTTCCCAACAAACCGTAAATG 1732
Qy 1547 CATTAATTTTAGTGAAAAATACTTTTCTAGCTTTTCAAAATTTTACGTCGATGAGAAACAGGA 1606
Db 1733 CAAAAACTGGTAAAAAATACTATTCTAGCTTTTAAAGGTCCTTAGAAAAAGAAAGACAGGA 1792

Qy 1607 TGTATTATTTTCATCTGGTGGCGGTTGCTGTGCCCTTCTTTTACATCGGAATATTT 1666
Db 1793 TGTATTATTTTCATCTGGTGGCGGTTGCTGTGCCCTTCTTTTATATTTGGAATTTAT 1852
Qy 1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGATCGAGTTAAATCTACATTAACCTGG 1736
Db 1853 TGGTTGTAAGACCGTTTACATAGAGTTTTCGACAGGATGATTAACCACTTTGACAGG 1912
Qy 1727 AAAAAGTATTTATCCCGTAAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGT 1786
Db 1913 AAAATTAGTGATCTCTGTAAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGT 1972
Qy 1787 ATATCTTAATCTATTAACTTTGGGAGTATTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGT 1846
Db 1973 TTAATCTTAAGCAATTAATTTAGGAGGAATTTTATTTGTTTCAGTGGGAAGAAATGAAGAAGT 2032
Qy 1847 CATGAACAACAGTTTATATCGATTCATAAAGAGATTTGATTTTATGAAGAAATGAAGAAGT 1906
Db 2033 CATGAACAGCAGTTTCAACCGTCTTATTAAGAGTTGATTAATTAAGAGGACAGGTGCT 2092
Qy 1907 ATAAACCGACGAAATATTTTATTAACAGGATTTCTGACTATATTTCCAGATATTTGCAAG 1966
Db 2093 ATTGATCAAGAGTGTTCATTTCAACCGGTTACTCAGACTTTGAACTTCAGATTTGTCAG 2152
Qy 1967 TATAAAAAATTTCTCAGTTTCAAGAAATGAAGAAATGAAGAAATGAAGAAATGAAGAAAT 2026
Db 2153 TGGTCAAAATTTCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2212
Qy 2027 ATTGGCCACGAGG 2040
Db 2213 ATCACATGCGCG 2226

RESULT 15
US-10-804-408-172
; Sequence 172, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: PANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 172
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-172

Query Match 9.3%; Score 647.6; DB 18; Length 2217;
Best Local Similarity 59.5%; Pred. No. 1.2e-81;
Matches 1222; Conservative 0; Mismatches 794; Indels 38; Gaps 6;

Qy 16 TGGCAATTATTTGATATGATAGTAGCAGTTGCAATTTCTGCAATCTTAACCAAGTCAATATACCAA 75
Db 173 TAGCGATGATTTCAACAGTTGTTGTTTATTTTCTGCAAGTTTGACATTAACATTAATTA 232
Qy 76 ATGCTGATTTAAATCG---TTCTGGAATTTTATATGATAGTAGAGTTAACTGATAGTTGAA 132
Db 233 CTCCCAATTTTAAAGCAATAAAGATTTATTTGTTGTTCTATGATACATTTATTTGCT 292
Qy 133 TTTTATATCTCGATGCGAGTTGATTTGATAGTAGAGTTAACTGATAGTTGAA 192
Db 293 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGCTATCTTTGAAGAGTTTAA 352
Qy 193 AAACATTTAACTAGTATATATTTTGAATTTTCTTACGGCAGTATCATTTTGTGTTG 252

Db 353 TGGTATTGAAATACAGCTTTTACTATATTTTCAATCAAGTTCAATATTTTATTTTATTTTA 412
Qy 253 AGAATAAATTTTCGCACTTTTCAAGACGTGGTGGTGTATTTTCAATTAATAAATTCGTTT 312
Db 413 AAAAATCAATTTACACACGACGCTTTTCTTTTCTTTTATTTGCTATGATTCGATTT 472
Qy 313 TGGTATACCTATTTAAACGTAATTAATTAAGCAGTTTAAAGGATAGCTTTCTATTTTCGACA 372
Db 473 TATTGTATCTATTGAATTTCAATTTTAAATAATATATCGAAATAATTTCTTACGCTAAGTTT 532
Qy 373 TCATCAAAAAGACGATTTCTAATTAACAGGCTGAACGATGGAATAATCAAGTTT 432
Db 533 CACGAGATACCAAGTTGTTTGTATAAGTAAGGATTTCTTTATCAAAAATGACCTTTA 592
Qy 433 TATTGTAATCACATAAACAATTTCAAAAATCTTTGTTGATTTGGTAGTTTGTAGTACAG 492
Db 593 GGAATAAT---ACGACCAATAATTTATTTGCTGTCTGCATCTTGGACTCTCTTGAAGG 649
Qy 493 AAATAGATAAATTTAATTTATCATTCAGCTCTATTTATTTCTGTGGAAGAGCTATAGAGT 552
Db 650 ATTGTTATGATTTGAAACATAAATCTGTTAAGGATAATAAACAAGATGCTCTTACTTCA 709
Qy 553 TTTCAACAGGAGTGGTGGACCACTCTTTTAAATCTTACCAAGTGGAGTTTGTAG--- 609
Db 710 AGTTAACTCTGTTAACTGTTGATCAAGCTTTTATTAACATACCCATTTGAATTTTGGTA 769
Qy 610 -----ACGTAAAGCAATTTGTTTTCAGATTTTGGTTGTTAGGTATTTGATGTAAGCGTTG 663
Db 770 AATACCAATACAGATTTATTAATGACATTTGAAGCAATGGGAGTGTATGTCATGTTA 829
Qy 664 ATATTAATTTCAATTCGGTTTACTCGCTTTGAAAAA CAAAAAATCCAACTGCTAGGTGACC 723
Db 830 ATGTAGAGGCACTTTAGCTTTGATAATATAGGAGAAAGCGAATCCAACTTTTGAAGGAT 889
Qy 724 ATAGCATTTGAATTTTTCACAAATTTTATAGCTTATAGCTCATATCATGATGAACGAC 783
Db 890 ATAGTGTATTACATATTTCTATGAATTTCTATAAATATAGTACCTTTATAGCAAAACGAT 949
Qy 784 TTTTGGATATACTCGGAGCGGTAGTCGGTTTAAATTTATTTGTTGGTATAGTTCTATTTGT 843
Db 950 TTTTGGATATACCGGTTGCTATTTATAGTTTTCATATAGTGGCATTTGTGCAATTTTTC 1009
Qy 844 TAGTTCCAAATTTATTCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAAAGTGGAC 903
Db 1010 TAGTTCCAAATTCAGAAAAGATGTTGGACCGGCTATCTTTTCTCAAAATAGAGTAGTGC 1069
Qy 904 AGAATGACGCAATTTTACATTTTACAAGTTTTCATCGATGATGTTGATGCTGAGGAGC 963
Db 1070 GTATGTTAGGATTTTGTAGATTTCTATAAATTCAGATCAATGCGAGTAGATGACAGAACAA 1129
Qy 964 GCAAAAAGACTTGTCTCAGCCAAACAGATGCAAGGTTGGTATGTTTAAAAATGGAA 1023
Db 1130 TTAAGAAGATTTTATTTAGTTTCAATCAATTAAGCAGGG-----CTAATGTTTAAAGTTAGA 1184
Qy 1024 AAACGATCTAGATTTACTTCAATTTGACATTTTATACGCAAAAAACAAGTTTAGACGAGT 1083
Db 1185 CGATGATCTTAGAATTTACTAAATAGGAAATTTATTCG-AAAAACAAGCATAGATGAGT 1243
Qy 1084 TACCAGTTTATTAATTTTAAATGCGGATATGAGTTAGTTGGTATGATGCTGACCTCA 1143
Db 1244 TGCCTCAATTTCTAATTTTAAAGGTGATATGAGTTTAGTAGGAACACGCGCTCCCA 1303
Qy 1144 CAGTTGATGATTTGAAAAATATATCTCTGGTCAAAAGACGATTTGAGTTTAAAAACAG 1203
Db 1304 CAGTTGATGATTAAGAAAGTATTAATTTCAACGCAAGACGCGCTTAGTTTAAAGCCAG 1363
Qy 1204 GGATTTACAGTCTCTGCGAGTTAGTGTCTGCTAGTAAATATCACAGACTTCCAGCAGCTAG 1263
Db 1364 GAATCAGTGTGTTGCGCAATATCTGTTAGAAATATATACATGATTTGATGAAATCG 1423
Qy 1264 TTCGTTGGACTTAGCATAATTTGATAATTTGGACTATCTGTTGAGATATTTAAATTTTAT 1323

Db 1424 TAAAGTTAGATGTTCAATATATCAATGATGTTCTATTTTGGTCAGATATTAAGATTATTC 1483
Qy 1324 TAAAGACAGTGAAAGTTGTTATGTTGAGAGGGAAGTAAGTAAAGTAT----- 1373
Db 1484 TCCTAACACACTAAAGGTAGTCTTACTTGGGACAGGTGCTAAGTAAAGGTAAAGTTGAAAG 1543
Qy 1374 -----ATGAAAGCTTTTGGTTCGGTCTTTCAGGGGACATTTGACTCAGTGTATTT 1426
Db 1544 GAATATATGAAATTTGTTCTGGTGGTCAAGTGGTGGTCACTAGCACACITGAACTT 1603
Qy 1427 GTTAAACCGTTTGGGAAGGAAGAACGTTTGGGTTAACTTTGATGAAGAGATGC 1486
Db 1604 TTTGAAACCCATTGGGAAAGAGATAGTTTGGTGAACCTTTGATGAAGAGATGC 1663
Qy 1487 AAGAAGCTTTTGAAGAAATGAAAAATGATACCAATTTTCTTCCAAACAAATCGCAATCT 1546
Db 1664 TAGGAGTATTTCTAAGAGAGAGATGTTATATCAATGCTCTTCTTCCAAACAAACCGTAATGT 1723
Qy 1547 CATTAATTTAGTGAATAATCTTTCTTAGCTTTTCABAAATTTAGCTGATGAGAAACCGA 1606
Db 1724 CAAAACCTTGTGAAAAAATACTATTTAGCTTTTAAAGTCTTGAAGAAAGAAAGACCGA 1783
Qy 1607 TGTATATTTCACTCGTGGCGCGTTGCTGTCCCTTCTTTTACATCGGAAACATTT 1666
Db 1784 TGTATCATATCATCTGTTGCGCTGTAGCAGTACCATTTCTTTATATGTTAAGTTATTT 1843
Qy 1667 TGGAGCAAGACGATTTATATGTAAGTATTTGATTCGAGTTTAATAATCTACATTTAAGTGG 1726
Db 1844 TGGTTGTAAGACCGTTTATATAGAGGTTTTCGACAGGATAGATAAACCAACTTTGACAGG 1903
Qy 1727 AAAACTAGTTATCCCGTAACAGATATTTTATTTGTTTCAGTGGGAAGAAATGAGAGGT 1786
Db 1904 AAAATTAGTGTATCTGTAACAGATAAATTTATTTGTTGAGTGGGAAGAAATGAAAAAGT 1963
Qy 1787 ATATCCTAAATCTATTAACCTTGGGAGTATTTTAAATGATTTTGTAAACAGTAGGAAT 1846
Db 1964 TTATCCTAAGCAATTAATTTAGAGGAAATTTTAAATGATTTTGTACAGTGGGACCA 2023
Qy 1847 CATGAACAAAGTTTAATTCGATGATGATAAGAGATTTGATTTATTTGAAAAAATGGAAGT 1906
Db 2024 CATGAACAGCAGTTTCAACCGTCTTATTAAGAAAGTTGATAGATTAAAGGACAGGTGCT 2083
Qy 1907 ATACCGACGAAATATTTATTAACAGCAGATATTTCTGACTATATTCAGATATTTGCAAG 1966
Db 2084 ATTGATCAAGAAGTGTTCATTTCAACCGGTTACTCAGACTTTGAACCTCGAATTTGTCAG 2143
Qy 1967 TATAAAAAATTTCTCAGTTTACAAAAGAAATGGAACAATATATTAACAATCAGAAGTAGTT 2026
Db 2144 TGGTCAAAATTTCTCATATGATGATGATGAACTTTTACATGAAGAAGCTGAGATTGTT 2203
Qy 2027 ATTTGCCACGAGG 2040
Db 2204 ATCACACATGCGG 2217

RESULT 16

US-10-804-408-168

; Sequence 168, Application US/10804408

; Publication No. US20040253617A1

; GENERAL INFORMATION:

; APPLICANT: FANRONG, KONG

; APPLICANT: GILBERT, GWENDOLYN

; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI

; FILE REFERENCE: 675002-2001

; CURRENT APPLICATION NUMBER: US/10/804,408

; PRIOR FILING DATE: 2004-03-19

; PRIOR APPLICATION NUMBER: PCT/AU02/01281

; PRIOR FILING DATE: 2002-09-18

; PRIOR APPLICATION NUMBER: AU PR 7749

; PRIOR FILING DATE: 2001-09-19

; NUMBER OF SEQ ID NOS: 182

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 168

; LENGTH: 2226
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
; US-10-804-408-168

Query Match 9.3%; Score 647.6; DB 18; Length 2226;
Best Local Similarity 59.5%; Pred. No. 1.2e-81;

Matches 1222; Conservative 0; Mismatches 794; Indels 38; Gaps 6;

Qy 16 TGGCATTATTTGATATGATAGCAGTTTGCATTTTCTCAATCTTTTAAACAAGTCATATACCAA 75
Db 182 TAGCGATGATTCACACAGTTTGTGTTTATTTTCTGCAAGTTTGCACATTAACATTAATTA 241
Qy 76 ATGCTGATTTAATTCG---TTCTGGAAATTTTATCATATATGATGGTTCATTTTTCAT 132
Db 242 CTCCTCAATTTAAAGCAATAAAGATTTATTTGTTCTTCTTATGATACATTTATTTGTCT 301
Qy 133 TTTTATATCTCGTATGCCAGTTTGAATTTGAGTATAGAGTAATCTGATAGAGTTTGA 192
Db 302 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCTGGCTATCTTGAAGAGTTTAA 361
Qy 193 AAACATTTAATCTAGTATATAATTTTGCATTTTCTTACGGCAGTATCATTTTGTGG 252
Db 362 TGGTATTTGAAATACAGCTTTTACTATATTTTTCATATCAAGTTTCATTTTATTTT 421
Qy 253 AGAATAATTTTCGACCTTTCAAGAGCTGGTGGCGGTGATTTTTCACATTAATAAATCTGTTT 312
Db 422 AAAACTCAITTTACACGACACGACTTTTCTCTTTTCTTTTATTTGCTATGAAATTCGATTT 481
Qy 313 TGGTATACCTTATTTAAGTAAATTTTAAAGCAGTTTAAAGGATAGCTTTCTATTTTTCGACAA 372
Db 482 TATTGATCTATTTGAATTCATTTTAAATAATTTTAAATAATTTTTCGAAATAATTTCTAGCTAAGTTT 541
Qy 373 TCTATCAAAAAAGACGATTTCTAATTTACACCGCTGACGATGGAAGAAATATGCAAGTTT 432
Db 542 CACGAGATACCAAGTTTGTGATTAACGAATAAGGATTTCTTTATCAAAAATGACCTTTA 601
Qy 433 TATTGAAATCAATAAACAATTTCAAAAAATCTTGTGCAATTTGTTAGTTAGTTAGTACAG 492
Db 602 GGAATAAAT---ACGACCATTAATATATCGCTCTGTCATCTTTGGACTCTCTGAAAGG 658
Qy 493 AAATAGATAAATTAATTTATCATTTACCTCGCTCTATTTATTTCTGTTGGAAGAGCTATAGAGT 552
Db 659 ATTGTTATGATTTGAAACATACTCGTTAAGGATATAATAACAAGATGCTCTTACTTTCAG 718
Qy 553 TTTTCAACAAGGGAAGTGGTCGACCGCTCTTTATAAATCTTACCAAGTGAAGTTTGTAG--- 609
Db 719 AGTTAACCTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCCATTGAAATTTTGGTA 778
Qy 610 -----ACGTAAAGCAATTCGTTTCAGATTTTTCAGATTTTTCAGTTTTCAGTTAGTGAAGCGTTG 663
Db 779 AATACCAAAATACAGATATTTAATTAATGACATTTGAAGCAATGGGAGTGAATTTGCAATGTTA 838
Qy 664 ATATTAATTCATTCGGTTTTCCTGCGTTGAAAAACAAAAATCCCAACTCTAGGTGAC 723
Db 839 ATGTAGGCACTTAGCTTTGATATAATAGGAGAAAGCGAATCCAACTTTTGAAGAT 898
Qy 724 ATAGCATTTGTAATTTTTCACAAAATTTTATAAGCCTTAGTCATATCATGATGAACGAC 783
Db 899 ATAGTGTATTACATATTTCTGAAATTTCTATAAATATAGTCACTTTATAGCAAAACGAT 958
Qy 784 TTTTGGATATCTCGGAGCGGTAGTCGGGTAAATTTATTTGTTGTTAGTTAGTTTCTATTTTGT 843
Db 959 TTTTGGATATCACGGTGTCTTATTAAGTTTGTCTATGTCATGTCGCAATTTTGTGCAATTTTTC 1018
Qy 844 TAGTTCCAAATTTTCGTAGAGATGGTGGACCGGCTATTTTGTCTCAGAAAACGAGTTTCGAC 903
Db 1019 TAGTTCCAAATTCAGAAAAGATGGTGGACCGGCTATTTTCTCAAAATAGAGTAGGTC 1078
Qy 904 AGAATGACGCAATTTTACATTTCTACAAGTTTTCGATCGATGATGTTGATGCTGAGGAGC 963
Db 1079 GTAATGTTAGATTTTATAGATTCTATAAATTCAGATCAATGCGAGTAGATGCAAGACAA 1138

779 AATACCAATACAGATATATTAATGACATTAAGCAATGGAGTGAATGTCATGTTA 838
Db
664 ATATTAAATTCATTCGGTTTACTCGGTTGAAACAAAAATCAATCTGCTAGTGACC 723
Qy
839 ATGTAGAGGACACTTAGCTTGTATATATAGAGAAAAAGCGAATCCAAACTTTTGAAGAT 898
Db
724 ATAGCATTTGAACCTTTTCCACAAATTTTATAGCCCTAGTCATATCATGATGAACGAC 783
Qy
899 ATAGTGTTATACATATTCATGAAATTCATATAATATAGTCACCTTATAGCAAAAGAT 958
Db
784 TTTTGGATATACCTCGAGCGGTAGTCGGGTTAAATTTTGTGGTATAGTTTCTATTGTTG 843
Qy
959 TTTTGGATATACCGGTGCTATATAGTTTGTCTCATATGTGGCATTTGTGCAATTTTTC 1018
Db
844 TAGTTCCAAATTTTCGTAGAGATGGTGGACGGGCTATTTTGTCTCAGAAACGAGTTGGAC 903
Qy
1019 TAGTTCCGCAATCAGAAAAAGATGGTGGACGGGCTATCTTTCTCAAAATAGAGTAGGTC 1078
Db
904 AGAATGGACGCATATTTACATTTCTACAAGTTTCGATCGATGTATGTTGATGCTGAGGAGC 963
Qy
1079 GTAATGGTAGGATTTTGTAGTTTCTTATAAATTCAGATCAATGCGAGTAGATGCGAAACAA 1138
Db
964 GCAAAAAGACTTCTCAGCCAAACCCAGATGCAAGGGTGGGTATGTTTAAATGGAA 1023
Qy
1139 TTAAGAAAGATTTATAGTTTACATCAATGACAGGG-----CTAATGTTAAGTTAGA 1193
Db
1024 AAACGATCCTAGAAATTTACTCCAAATGGACATTTTCATACGCAAAACAAAGTTTGAAGAGT 1083
Qy
1194 AGATGATCCTAGAAATTTACTAAAAATAGGAAAAATTTATTCG-AAAAACAAAGCATAGATGAT 1252
Db
1084 TACCACAGTTTATAAATTTTAAATTTGCGCATAGCTAGTTGGTACACGCTCCACTA 1143
Qy
1253 TGCCTCAATTTCTAATGTTTTTAAAGGTGATATGATGTTTAGTAGGAAACACGCGCTCCCA 1312
Db
1144 CAGTTGATGAATTTGAAAAATATACTCTGCTCAAAAGAGACGATTCAGTTTAAACAG 1203
Qy
1313 CAGTTGATGAATATGAAAGTATTAATTCACGCAAGAGCGAGCTTAGTTTAAAGCAG 1372
Db
1204 GGATTACAGGTCCTGCGAGGTTAGTGGTGTAGTATATATACAGACTTCGACGACGTAG 1263
Qy
1373 GAATCACTGGTTTGTGCAAAATATCTGGTAGAAATAATATTAAGTTTGTGATGAATCG 1432
Db
1264 TTCGGTTGACCTTAGCATACATGATTAATTTGGACTATCTGGTCAGATATTAATTTTAT 1323
Qy
1433 TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTAAGATATTC 1492
Db
1324 TAAAGACAGTGAAAGTTGTATTTGTTGAGAGAGGGAAGTAAGTAAAGTAT----- 1373
Qy
1493 TCCTAACTAAAGGTAGTTTACTCGGACAGAGCTNAGTAAGGTAGGTTTGAAG 1552
Db
1374 -----ATGAAAGTTTGTGGTGGTTCTTCAGGGGACATTTGACTCACTTGTATTT 1426
Qy
1553 GAATATATGAAAAATTTGTCTGGTTGGTTCAAGTGGTGGTCTAGCAGACTTGAACCT 1612
Db
1427 GTTAAACCGTTTGGAGGAGAGAACGTTTGGTGGTAACTTTGATTAAGAGGATGC 1486
Qy
1613 TTTGAAACCCATTTGGGAAAAAGAGATAGGTTTGGGTAACTTTGATTAAGAGATGC 1672
Db
1487 AAGAAGTCTTTGAGAAATGAAAAATGTATCCATGTTTACTTTCCAAACAAATCGCAATCT 1546
Qy
1673 TAGGAGTATCTAGAGAGAGATTTGATATCATTTGCTTTCTTCCAAACAAACCGTATGT 1732
Db
1547 CATTAATTTAGTGAATAATCTTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACAGA 1606
Qy
1733 CAAAAACTTGTAATAAATACTATTCCTAGCTTTTAAAGTCTCTTAAAGAAAGAACACAGA 1792
Db
1607 TGTTATTTATTCATCTGGTGGCGGTTGCTGTCCTCTTTTACATCGGAAACACTTT 1666
Qy
1793 TGTTATCATATCATCTGGTGGCGGTGAGCAGTACCATTTCTTTTATATTTGGTAGTTATT 1852
Db
1667 TGGAGCAAGACGATTTATATTTGAAGTATTTGAATCGAGTTTAAATAATCTACATTAACCTGG 1726
Qy
1853 TGGTTGTGAAGACCGTTTATATATAGAGTTTTCGACAGGATAGATAAACCACTTTGACAGG 1912
Db

RESULT 18

US-10-804-408-164
; Sequence 164, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 164
; LENGTH: 2217
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-164

Query Match 9.2%; Score 646; DB 18; Length 2217;
Best Local Similarity 59.4%; Pred. No. 2e-81;
Matches 1221; Conservative 0; Mismatches 795; Indels 38; Gaps 6;

Qy 16 TGGCATTATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAAACAAGTCATATACCA 75
Db 173 TAGCGATGATTAACCGTTTGGTTTATTTTCTGCAAGTTTGACATTAACATTAATTA 232
Qy 76 ATGCTGATTTAAATCG---TTCTGGAAATTTTATCATATGATGGTTTCATTATTTGTCAT 132
Db 233 CTCCCACTTTAAAGCAATTAAGATTTATTTGTTGTTCTTATGATACATTTATTTGCT 292
Qy 133 TTTTATATCTCGTATGCCAGTTGAAATTTGAGTATAGAGTAATCTGATAGAGTTGAAA 192
Db 293 TTTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTCGCTATCTTGAAGAGTTTAAAA 352
Qy 193 AAACATTTAACTAGTATATAATTTTGCATTTTCTTACGGCAGTATCATTTTGTGG 252
Db 353 TGGTATGAAATACAGCTTTTACTATATTTTCAATCAAGTTTCATTATTTTATTTTAA 412
Qy 253 AGAATAATTTCCGACCTTTCAAGACGTTGGTGGCGGTGATTTTCACTTAATAAACCTCGTT 312
Db 413 AAAAAGCTCAATTACACGACGACGACTTTTCTTTTACTTTTATTTGCTATGAATTCGATTT 472

| | | | |
|------------------------------------------------------------------|------|-----------------------------------------------------------------|------|
| Matches 1220; Conservative 0; Mismatches 796; Indels 38; Gaps 6; | | | |
| Qy | 16 | TGGCAATTTATGATATAGAGTGGCAATTTCTGCAATCTTAAACAAGTCATATACCAA | 75 |
| Db | 173 | TAGCGATGATTCAAACAGTTGTGTTATGTTCTGTAAGTTTGCACATTAACATTAATCA | 232 |
| Qy | 76 | ATGCTGATTTAAATCG--TTCTGGAATTTTATCAATAATGATGGTTTCATTTTTCAT | 132 |
| Db | 233 | CTCCCAATTTAAAGCAATAAAGATTTATTTGTTGTTCTATTGATACATATATTGTCT | 292 |
| Qy | 133 | TTTTTATATCTCGTATGCCAGTTGAAATTGAGTATAGAGTAAATCTCATAGAGTTTGAAA | 192 |
| Db | 293 | TTTATCTTTCTGATTTTACAGAGCTTTTGGAGTCGTGCTATCTTGAAGAGTTTAAAA | 352 |
| Qy | 193 | AAACATTTAACTATGATPATAAATTTTGCATTTTCTTAACGGCAGTATCAATTTTGTGG | 252 |
| Db | 353 | TGGTATTTGAAATACAGCTTTTACTATATTTTCAATCAAGTTTCATTTATTTTATTTT | 412 |
| Qy | 253 | AGATTAATTTTCGACTTTCAAGACGTGGTCCGGTATTTTCAATTAATAAATCTTCGTTT | 312 |
| Db | 413 | AAAACTCTTTTACAAACAGCAGCTTTTCCTTTTACTTTTATTTGCTATGAATTCGATTT | 472 |
| Qy | 313 | TGGTATACCTATTAAACGTAAATTAAGCAGTTTAAAGGATAGCTTTCTATTTTCGACAA | 372 |
| Db | 473 | TATTAATCTATTGAAATTCATTTTAA--AAATTAATCGNAATATCTTACGCTAAGT | 529 |
| Qy | 373 | TCTATCAAAAAGACGATTTCTAATTAACAAGCTGAAACGATGGGAAATATGCAAGTTT | 432 |
| Db | 530 | TTTCACGAGATACCAAGTTGTTTGTATACGAAATAAGGATTTCTTTTCAAAAATGACCT | 589 |
| Qy | 433 | TATTTGATCACATAAACAATTCAAAATAATCTTGTGTCATTTGGTAGTTTATAGTACAG | 492 |
| Db | 590 | TTAGGAACAAATACGACCATTAATATATCGCTGTCTGTATCTTGGACTCTCTGAAAAGG | 649 |
| Qy | 493 | AAATAGATAAAATTAATTTATCAATTAACGCTCTATTATCTGTGGAAGAAGCTATAGAT | 552 |
| Db | 650 | ATTGTTATGATTGGAACATAACTCGTTAAGGATAATAACAAGATGCTCTTACTTCAG | 709 |
| Qy | 553 | TTTCAACAAGGAAGTGGTGCACACGCTCTTTATPAATCTAACCAAGTGAGTTTATAG-- | 609 |
| Db | 710 | AGTTAACTGCTTAACCTGTTGATCAAGCTTTTATTAACATACCCATTGAAATTTTGGTA | 769 |
| Qy | 610 | -----ACGTAAACCAATTCGTTTCAGATTTTGGATTTGTTAGTTATGATGAAGCGTTG | 663 |
| Db | 770 | AATACCAAAATCAAGATATTATTAATTGACATTTGAAGCAATGGGAGTGATTTGTCAATGTTA | 829 |
| Qy | 664 | ATATTAATTCATTCCGTTTACTCGCTTGAACCAACAAAATCCAACTGCTAGTGACCC | 723 |
| Db | 830 | ATGTAGAGGCACTTAGCTTTGATTAATATAGGAGAAAGCGAATCCAACTTTTGAAGAT | 889 |
| Qy | 724 | ATAGCATTTGTAACCTTTTCCCAAATTTTATAAGCCCTAGTCATATCATGATGAACGAC | 783 |
| Db | 890 | ATAGTGTATTACATATCTATGAAATTTCTATAAATATAGTCACCTTATAGCAAAACGAT | 949 |
| Qy | 784 | TTTTGGATATCTCGAGCGGTAGTCGGTTTAAATTTTGTGGTATAGTTTCTATTTTGT | 843 |
| Db | 950 | TTTTGGATATCATGGGTGCTATTATAGTTTGTCTCATATGGCAATTTGTGCAATTTTTC | 1009 |
| Qy | 844 | TAGTTTCAAATTTTCTGATAGATGTTGACCGGCTATTTTGTCTCAGAAAACGAGTTCGAC | 903 |
| Db | 1010 | TAGTTTCCGAATCAGAAAAGATGGCGACCGGCTATCTTTCTCAAATATAGATAGTCT | 1069 |
| Qy | 904 | AGAAATGGACGATATTTACATTTCTACAAGTTTTCGATCGATGATGTTGATGCTGAGGAC | 963 |
| Db | 1070 | GTAATGGTAGGATTTTATGATTTCTATAAATTCAGATCAATGCGAGTAGTCGAAACAAA | 1129 |
| Qy | 964 | GCAAAAAGACTTCTCAGCCAAAACCGATGCGAGGGTGGGTATGTTTAAATGGGAA | 1023 |
| Db | 1130 | TTAAGAAAGATTTTATTAAGTTTCACAATCAAAATGACAGGG-----CTAATGTTTAA | 1184 |
| Qy | 1024 | AAACGATCCTAGAATTTACTCCAAATTTGGACATTTTCATACGCCAAAACCAAGTTTACGAGT | 1083 |
| Db | 1185 | AGATGATCCTAGAATTTACTAAATATAGGAAATTTTATTCG-AAAAACAAGCATAGATGAAT | 1243 |

| | | | |
|----|------|--------------------------------------------------------------------|------|
| Qy | 1084 | TACCACAGTTTATATATGTTTAAATTTGGCGATATGAGTCTAGTTGGTACACGTCACCTA | 1143 |
| Db | 1244 | TGCTCTCAATTTCTATATATGTTTAAAAAGGTGATATGAGTTTGTAGTGAACACGCGCTCCA | 1303 |
| Qy | 1144 | CAGTTGATCAATTTTGAATAATATACCTCTGGTCAAAAGAGACGATTTGAGTTTAAACCAG | 1203 |
| Db | 1304 | CAGTTGATGAATATGAATAAGATTAATCAACGCAAGAGCGACGCTTAGTTTAAAGCCAG | 1363 |
| Qy | 1204 | GGATTAACAGGCTCTCTCGCAGGTTAGTGGTCGTAGTAAATATACAGAGCTTTCCGACGCTAG | 1263 |
| Db | 1364 | GAATCACTGGTTTGTGCAAAATATCTGCTAGAAATAATATTAATCTGATTTGATGAATCG | 1423 |
| Qy | 1264 | TTCCGTTGGAACCTTAGCATACATTTGATPAATTTGGATCTATCTGGTCAGATATTTAAATTTTAT | 1323 |
| Db | 1424 | TAAAGTTAGATGTTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTTAAGATTTATC | 1483 |
| Qy | 1324 | TAAAGACAGTGAAGGTTGTATTTGTTTCAGAGAGGGAAGTAGTAAAGATAT----- | 1373 |
| Db | 1484 | TCATAACACTAAAGGTAGTTTCTCGGACAGGAGCTAAGTAAAGGTAAAGTTTGAAG | 1543 |
| Qy | 1374 | -----ATCAAAAGTTTGTGTTGGTCGGTCTTTCAGGGGGACATTTTGACTCAGCTTGATTTT | 1426 |
| Db | 1544 | GAATAAATGAATAATTTGCTGGTTCAAGTGGTGGTCACCTAGCACACTTGAACCT | 1603 |
| Qy | 1427 | GTTAAAAACCGTTTGGAGGAAGAAACGTTTTTGGGTAAACATTTTGAATAAAGAGATGC | 1486 |
| Db | 1604 | TTTGAACCCATTTGGGAAAAAGAGATAGTTTGGGTAAACCTTTGATAAAGAAGATGC | 1563 |
| Qy | 1487 | AGAACTCTTTGAAGAATAAATAATGATCCATGTTACTTTTCCCAACCAATCGCAATCT | 1546 |
| Db | 1664 | TAGGATGATTTCTAAGAGAAGATTTGTAATATCATTTGCTTTTCCCAACCAACCGTAAATG | 1723 |
| Qy | 1547 | CATTAATTTAGTGAAAAATACCTTTCTTAGCTTTTCAAAATTTTACGTGATGAGAAACCGA | 1606 |
| Db | 1724 | CRAAACTTGGTAAAAAATACATTTCTAGCTTTTAAAGTCTTTAGAAAAGAGAACGAGA | 1783 |
| Qy | 1607 | TGTTATTTTCACTGGTGGCGGCTGCTGCTCCCTCTTTTATCATCGGAAAACTATT | 1666 |
| Db | 1784 | TGTTATCATATCATCTGGTCCGCTGTAGCAGTACCATTCTTTTATATTTGTTAAGTTTAT | 1843 |
| Qy | 1667 | TGGAGCAAGACGATTTATTTGAAGTATTTGATCCAGTTTAAATAATCTCATTTAACTGG | 1726 |
| Db | 1844 | TGTTTGAAGACCGTTTATATAGAGGTTTTCGACAGATAGATAAAACCACTTTGACAGG | 1903 |
| Qy | 1727 | AAAACTAGTTTATCCCGTAAACAGATATTTTATTTGTTTCACTGGGAGAAATGAAGAAGT | 1786 |
| Db | 1904 | AAATTAGTGTATCTGTAAACAGATAAATTTATTTGTTCACTGGGAGAAATGAAGAAAT | 1963 |
| Qy | 1787 | ATATCCTAAATCTATTAACCTGGGAGATTTTTTTAATGATTTTGTAAACAGTAGGAAT | 1846 |
| Db | 1964 | TTATCCTAAGGCAATTTAAATTTAGGAGGAAATTTTAAATGATTTTGTACAGTGGGACA | 2023 |
| Qy | 1847 | CATGAACAACAGTTTAAATCGATTTGAAGAAGATTTGATTTTAAAGAAATTTGAAGT | 1906 |
| Db | 2024 | CATGAACAGCAGTTTCAACCGCTCTTATTAAGAAAGTTGATAGATTAAAGGAGCAGGTGCT | 2083 |
| Qy | 1907 | ATAACCGACGAAATATTTTAAACACAGGATATTTCTGACTATATTTCCAGAAATTTGCAAG | 1966 |
| Db | 2084 | ATTGATCAGAGAGTGTTCATTCAAACGGGTATCTCAGACTTTGAACTCAGAAATTTGTCAG | 2143 |
| Qy | 1967 | TATAAAAAATTTCTCAGTTTCAAAAGAAATGGAAACAATATATTTAAACAAATCAGAGTAGTT | 2026 |
| Db | 2144 | TGGTCAAAATTTCTCTCATATGATGATGAACTCTTCAATGAAGAAGCTGAGATTGTT | 2203 |
| Qy | 2027 | ATTTGCCACGGAGG 2040 | |
| Db | 2204 | ATCACACATGGCGG 2217 | |


```

; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 170
; LENGTH: 2226
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
; US-10-804-408-170

Query Match          9.2%; Score 644.4; DB 18; Length 2226;
Best Local Similarity 59.4%; Pred. No. 3.4e-81;
Matches 1220; Conservative 0; Mismatches 796; Indels 38; Gaps 6;

Qy      16  TGGCAATTATTTGATATGATGACGAGTTGCAATTCCTGCAATCTTAAACAGTCATATACCAA 75
Db      182  TAGCGATGATTCAAACCGTTCTGGTTATATTTTCTGCAAGTTTGACATTAAACATTAATTA 241

Qy      76  ATGCTGATTTAAATCG---TTCCTGGAATTTTTTATCATAAATGATGGTTCAATTTTTCGAT 132
Db      242  CTCCCAACTTTTAAAGCAATAAAGATTTAATGTTTGTTCATGATACATATATATGTGCT 301

Qy      133  TTTTATATCTCGTATGCCAGTTGCAATTTGAGTATAGAGGTAATCTCGATAGAGTTTGAAA 192
Db      302  TTATCTTCTGATTTTACAGAGACTTTTGGAGTCGTGGGCTACTCTGAAGAGTTTAAAA 361

Qy      193  AAACATTTAATAGTATAGTATAATATTTGCAATTTTCTTACGGCAGTATCATTTTTGTGTG 252
Db      362  TGGTATTTGAAATACAGCTTTTACTATATATTTTATCATCAAGTTTCATTTATTTTATTTTA 421

Qy      253  AGATAATTTTCGACTTTTCAAGCGTGGTGGCGGTATTTTACACATTAAATAACTTCGTTT 312
Db      422  AAAACTCTTTTACACGACAGCACTTCTCTTTTACTTTTATTACTATGAATTCGATTT 481

Qy      313  TGGTATACCTATTTTAAAGTAATTTAAGCAGTTTAAAGTAGCTTCTATTTTCGACAA 372
Db      482  TATATATCTATTGAAATTCATTTTAAAAATATATCGAAATATCTTACGCTAAGTTT 541

Qy      373  TCTATCAAAAAAGACGATCTTAATTAACAACGGCTGAACGATGGGAAAAATATGCAAGTTT 432
Db      542  CACGAGATACCAAAGTTGTTTGTGATAACGAATAAGGATCTCTTATCAAAATGACCTTTA 601

Qy      433  TATTTGAATCACATAAACAATTTCAAAAAAATCTTTGTTGCAATTCGTAGTTTTAGGTACAG 492
Db      602  GGAATAAAT---ACGACCATAATTTATATCGCTGTCTGTATCTTGGATTCCTCTCGAAAAG 658

Qy      493  AAATAGATAAAATTTAATTTATCATTTACGCTCTATTTATTCGTGGAAGAAGCTATAGAGT 552
Db      659  ATTGTTATGATTGAAACATAACTCGTTAAGGATAATAAACAAGATGCTCTTACTTCAG 718

Qy      553  TTTCAACAAGGGAAGTGGTGCACACCGTCTTTATAAAATCTACCAAGTGAGTTTTTAG--- 609
Db      719  AGTTAACTGCTTAACTGTTGATCAAGCTTTTATTAACATACCCATTGAATTTATTTGTA 778

Qy      610  -----ACGTAAAGCAATTCGTTTTCAGATTTTGAATTTAGTTATTTAGTATGATAGCGTTG 663
Db      779  AATACCAAAATACAAGATATTTAATGACATTTGAAGCAATGGGAGTGATTTGCTCAATGTTA 838

Qy      664  ATATTAAATTCATTCGGTTTTTACTCGGTTGAAAAACAAAAATCCAACTGCTAGGTGACC 723
Db      839  ATGTAGAGGCACTTAGCTTTTGATTAATATAGAGAAAAACGGAATCCAAACTTTTGAAGAT 898

Qy      724  ATAGCATTTGAACCTTTTTTCCAAAAATTTTTTATAAGCCCTAGTTCATATCATGATGAAACGAC 783

```


Db 1973 TTATCCTAAGCAATTAATTTAGGAGGAAATTTTAAATGATTTTTTGTGCACAGTGGGACA 2032
Qy 1847 CATGAACAACAGTTTAAATCGATGATTAAGAGAGATTGATTTATGAAAAAATGGAAGT 1906
Db 2033 CATGAACAGCAGTTTCAACCGCTTTATTAAGAGAGTTGATAGATTAAAGGGACAGATGCT 2092
Qy 1907 ATACCCAGCAAAATATTTATTTCAACAGGATATTCTGACTATATTCCAGAAATTTGCAAG 1966
Db 2093 ATTGATCAGAAGTGTTCATTTCAACCGGTTACTCAGACTTTGAACTCGAAATTTGTGAC 2152
Qy 1967 TATAAAAAATTTCTCAGTTTCAAGAAATGGAACAATATATTAAACAAATCAGAAGTAGTT 2026
Db 2153 TGGTCAAAATTTCTCTCATATGATGATGAACTCTTTACATGAAAGAGCTGAGATGTTT 2212
Qy 2027 ATTTGCCACGGAG 2040
Db 2213 ATCACATGGCG 2226

RESULT 21
US-10-804-408-165
; Sequence 165, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 165
; LENGTH: 2225
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-165

Query Match 9.18; Score 638.8; DB 18; Length 2225;
Best Local Similarity 59.58; Pred. No. 2.1e-80;
Matches 1223; Conservative 0; Mismatches 792; Indels 39; Gaps 7;

Qy 16 TGGCAATTAATTTGATATGATAGCAGTTGCAATTTCTGCAATCTTAAACAGTCATATACCAA 75
Db 182 TAGCGATGATTCAAACAGTTGTGGTTTATTTTCTGCAAGTTTGACATTAACATTAATTA 241
Qy 76 ATGCTGATTTAAATCG---TTCGGAAATTTTATCATAAATGATGGTTTCATTTTTCGAT 132
Db 242 CTCCCAATTTTAAAGCAATAAAGATTAATTTGTTGTTCTATTGATACATTAATTTGTCT 301
Qy 133 TTTTATATCTCGTATCGCAGTTCAATTTGAGTATAGAGTAATCTCATAGAGTTTGAAA 192
Db 302 TTTAATCTTCGATTTTACAGACACTTTTGGAGCTGGTGGCTATCTTGAAGAGTTTAAAA 361
Qy 193 AAAACATTAACATATAGTATTAATTTGCAATTTTCTTACGGCAGTATCAATTTTGTGG 252
Db 362 TGGTATTTGAATAACAGCTTTTACTATATTTTCATATCAAGTTTCATTTTATTTTAA 421
Qy 253 AGAATAATTTGCACTTTCAAGAGCGTGGCCGTTATTTTCAATTAATTAATTAACCTCGTTT 312
Db 422 AAAACTCAATTTACAATGACACGACTTTCCCTTTTCTCTTTTATTTGCTATGAAATTCGATTT 481
Qy 313 TGGTATACCTATTTAAACGTAATTAATTAAGCAGTTTAAAGTATAGTATTTCTATTTCGCAA 372
Db 482 TATTATATCTATGAAATTCATTTTTTAAATATTAATCGAAATATTTCTTACGCTAAGTTT 541
Qy 373 TCTATCAAAAAAGACGATCTTAATTAACAACGGCTGAACGATGGGAAATATGCAAGTTT 432
Db 542 CACGAGATACCAAGTTGTTTGTGATAACGAATAAGGATCTTTATCAAAAAATGACCTTTA 601

Qy 433 TATTTGAATCACAATAACAATAAATCTTGTTCGATTTGGTAGTTTATAGGTACAG 492
Db 602 AGAATAAAT---ACGACCATAAATATATCGCTGTCTGTATCTTTGGACTCTCTGAAAGG 658
Qy 493 AAATAGATAAAATTAATTTATCATTAACCGCTCTATTATTCTGTGGGAAGACTATAGAGT 552
Db 659 ATTGTTATGATTTGAAACATAACTCGTTAAGGATATAAACAAGATGCTCTTACTTTCAG 718
Qy 553 TTTCAAACAAGGAGTGGTCGACCAGCTCTTTATAAATCTTACCAAGTAGTTTTTAG--- 609
Db 719 AGTTAACTGCTTAACTGTGTGATCAAGCTTTTATTAACTACCACTTGAATTAATTTGGTA 778
Qy 610 -----ACGTAAAGCAATTCGTTTCAGATTTTCAGTTTGTAGTATTGATGTAAGCGTTG 663
Db 779 AATACCAAAATACAAGATATTTAATTAATGACATTAAGCAATGGGAGTGTGTCAATGTTA 838
Qy 664 ATATTAAATTCATTCGGTTTTTACTCGCTTGAATAAACAATAAATCTTACCTAGGTGACC 723
Db 839 ATGTAGAGGCACTTAGCTTTTGATATATAGGAGAAAGCGAATCCAACTTTTGAAGGAT 898
Qy 724 ATAGCAATGTAACTTTTCCACAAATTTTATAAGCCTAGTCAATATCATGATCAAAACGAC 783
Db 899 ATAGTGTATTACATATTCTATGAAATTTCTATAAATATAGTCACTTATAGCAAAACGAT 958
Qy 784 TTTTGGATATACTCGGAGCGGTAGTCGGGTTAAATTAATTTTGTGTTAGTATTCTTATTGTT 843
Db 959 TTTTGGATATCATGGGTGCTATTATAGGTTTGTCTATATGTGGCATTTTGGCAATTTTC 1018
Qy 844 TAGTTCCAATTTATTCGTAGAGATGGTGGACCGCTATTTTGTCTCAGAAACGAGTTGGAC 903
Db 1019 TAGTTCGCAATCAGAAAGATGGTGGACCGCTATCTTTCTCAAAATAGAGTAGGTC 1078
Qy 904 AGAATGAGCGCATTTTACATTTCTACAAGTTTTCGATCGATGTTGTTGTCGAGGAGC 963
Db 1079 GTAATGTTAGGATTTTGTAGATT-TATAAATTCAGATCAATGCGAGTAGATGCAAGACAA 1137
Qy 964 GCAAAAAAGACTTGTCTCAGCAAAACAGATGCAAGGGTGGGTATGTTTTAAATGGGAA 1023
Db 1138 TTAAGAAAGATTTTATTTAGTTTCAAAATCAAAATGACAGG-----CTAATGTTAAGTTAGA 1192
Qy 1024 AAACGATCCTAGAAATTTACTCCAATTCGACATTTATACGCAAAACAAAGTTTACAGAGT 1083
Db 1193 CGATGATCCTAGAAATTTACTAAATAGAAATTTATTCG-AAAAACAGCATAGATGAGT 1251
Qy 1084 TACCAAGCTTTTATAATGTTTAAATTTGGCGATATGAGTCTAGTTGGTACACGCTCCACTA 1143
Db 1252 TGCCTCAATTTCTATAATGTTTAAAGGTGATATGAGTTTAGTAGGAACACGCGCTCCCA 1311
Qy 1144 CAGTTGATGAATTTGAAAAATATATCTCTGGTCAAAAGACAGATTTGAGTTTTTAAACCAG 1203
Db 1312 CAGTTGATGAATATGAAGAATGATTAATTCACGCAAGAGCGACGCTTAGTTTAAAGCAG 1371
Qy 1204 GGATTAACAGTCTCTCGCAGGTTTAGTGGTCTAGTAAATATACAGACTTCGACAGCTAG 1263
Db 1372 GAATCACTGTTTGTGGCAAAATATCTGGTAGAAATTAATTAATCTGATTTTGATGAATCG 1431
Qy 1264 TTCGTTTGGACTTTAGCATACATTTGATAATTTGGAATATCTGTTGTCAGATATTTAAATTTAT 1323
Db 1432 TAAAGTTAGATGTTCAATATATCAATGAATGGTCTATTTGGTCAGATATTAAGATTATTC 1491
Qy 1324 TAAAGACAGTGAAGTTGTTATTTGTTGCTCTTTCAGGGGACATTTGACTCAGTTGATTT 1373
Db 1492 TCCTAACTTAAAGGTAGTCTTACTTTGGGACAGGAGCTAAGTAAAGGTAGGTTTGAAG 1551
Qy 1374 -----ATGAAAGTTTGTGTTGCTCTTTCAGGGGACATTTGACTCAGTTGATTT 1426
Db 1552 GAATATAAGAAATTTGCTGTTGTTCAAGTGGTGGTCTCTAGCACCTTGAACCT 1611
Qy 1427 GTTAAACCGTTTGGAAAGGAAGAACGTTTTTGGGTTAACTTTTGGTAAAGAGGATGC 1486
Db 1612 TTTGAAACCCATTTGGGAAAAAGAGATAGGTTTTTGGGTTAACTTTGATAAAGAGATGC 1671

| | | | |
|----|------|---------------------------------------------------------------------------------------------------------------------------|------|
| Qy | 1487 | A G A A G T C T T T T G A A G A A T G A A A A A T G T A T C C A A T G T A C T T T C C A C A A A T C G C A A T C T | 1546 |
| Db | 1672 | T A G G A G T A T T C T A A G A G A G A T T G T A T A T C A T T G C T T C T T T C C A C A A C G T A A T G T | 1731 |
| Qy | 1547 | C A T T A A T T T A G T G A A A A A T A C T T T C T T A G C T T T C A A A A T T T A C G T G A T C A G A A A C C A G A | 1606 |
| Db | 1732 | C A A A A C T T G G T A A A A A N T A C T A T T C T A G C T T T T A G G T C C T T A G A A A A G A A G A C C A G A | 1791 |
| Qy | 1607 | T G T T A T T A T T T C A T C T G G T C G G C G G T T G C T G C C C T T C T T T T A C A T C G G A A A A C T P A T T | 1666 |
| Db | 1792 | T G T T A T C A T A T C A T C T G G T C G C G T G T A G C A G T A C C A T T C T T T T A T A T T C G T A A G T T A T T | 1851 |
| Qy | 1667 | T G G A G C A A A C A C A G A T T T A T A I T G A A G T A T T T G A T C G A G T T A A T A A A T C T A C A T T A A C T G G | 1726 |
| Db | 1852 | T G S T T G T A A G A C C G T T T A T A T A G A G S G T T T C G A C A G G A T A G A T A A A C A A C T T T G A C A G G | 1911 |
| Qy | 1727 | A A A A C T A G T T T A T C C C G T A A C A G A T A T T T T T A T T G T T C A G T G G G A A A A A T A A G A A G T | 1786 |
| Db | 1912 | A A A A T T A G T G T A T C C T G T A A C A G A A A A T T A T T G T T C A G T G G G A A A A A T A A A A A A G T | 1971 |
| Qy | 1787 | A T A T C C T A A A T C V A T T A A C T T G G G G A G T A T T T T T A T G A T T T T T G T A A C A G T A G G A A C T | 1846 |
| Db | 1972 | T T A T C T C A G G C A A T T A A T T T A G G A G A A T T T T T A T G A T T T T T G T C A C A G T G G G A C A | 2031 |
| Qy | 1847 | C A T G A A C A A C A G T T T A A T C G A T T G A T A A A A G A G A T T G A T T A T T T G A A A A A A A A T A G A A G T | 1906 |
| Db | 2032 | C A T G A A C A G C A G T T C A A C C G T C T T A T T A A A A G A A G T T G A T A G A T T A A A A G G A C A G T G C T | 2091 |
| Qy | 1907 | A T A A C C G A C A A A T A T T A T T C A A A C A G G A T A T C T G A C T A T A T T C C A G A N A T A T T G C A A G | 1966 |
| Db | 2092 | A T T G A T C A A A A G T T T C A T T C A A A C G G G T T A C T C A G A C T T T G A A C C T C A G A A T T G T C A G | 2151 |
| Qy | 1967 | T A T A A A A A A T T T C T C A G T T A C A A A A A T G G A A C A A T A T A T T A A C A A A T C A G A A G T A G T T | 2026 |
| Db | 2152 | T G G T C A A A A T T T C T C A T A T A T G A T A T G A A C T C T T A C A T G A A A G A A G C T G A G A T T G T T | 2211 |
| Qy | 2027 | A T T T G C A C G G A G | 2040 |
| Db | 2212 | A T C A C A C A T G G C G | 2225 |

RESULT 22

```

RESUL1. 22
US-10-182-960-1
; Sequence 1, Application US/10182960
; Publication No. US20030186392A1
; GENERAL INFORMATION:
; APPLICANT: Tremby, Janine, et al.
; TITLE OF INVENTION: Biopolymer Thickner
; FILE REFERENCE: 64376
; CURRENT APPLICATION NUMBER: US/10/182,960
; CURRENT FILING DATE: 2002-08-03
; PRIOR APPLICATION NUMBER: PCT/US01/03404
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/241,098
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/179,888
; PRIOR FILING DATE: 2000-02-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 6850
; TYPE: DNA
; ORGANISM: Lactococcus lactis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (174)..(488)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (528)..(977)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1020)..(1796)
; FEATURE:

```


Db 4229 AAAATACCTTAGAAGAAACACAGATTGATTTATTTTCGAGTGTGCTGCGTAGCGGTT 4288
Qy 1641 CCCTTCTTTTACATCGGAAACTATTTTGGAGCAAGAGATTTATATTGAAGTATTGTAT 1700
Db 4289 CCTTTTCTTTGGTTAGGTAACATTTTCGGTGCAGAGACAGCTCTATATTGAAATATTGAC 4348
Qy 1701 CGAGTTAATAATCTACATAACTGGAAGAACTAGTTTATCCCGTAACAGATATTTTATT 1760
Db 4349 CGGATGATTAACCACTTAAACAGGAAATAGTTTATCCAGTTACTGATAAGTTTATA 4408
Qy 1761 GTTCAGTGGGAAGAAATCAAGAGGTATATCTTAAATCTATTAACTTGGGGAGTATTTT 1820
Db 4409 GTTCAATGGGAGAGTTTAAAGAAAGTTTACCTTAAACCAATTAATTTAGGAGNAITTC 4468
Qy 1821 TAATGATTTTGTAACTAGTAGGAATCATGAACCAAGTTTAACTGATGATTAAGAGAGA 1880
Db 4469 TAATGATTTTGTAACTAGTAGGAATCATGAACCAAGTTTAACTGATGATTAAGAGAGA 4528
Qy 1881 TTGATTTTGTAAAGAAATGGAAGTATTAACCGAGCAATATTTTATTTCAACAGGATTT 1940
Db 4529 TTGATGAACTTGTACGGGATGGTGAATCGAAGACGATGATTTATCATGCAAAATTTGGGTACT 4588
Qy 1941 CTGACTATATTTCCAGAAATATTGCAAGTATATAAAATTTCTCAGTTACAAAGAAATGGAAC 2000
Db 4589 CACTTATGAACCTTAATATATACTAAATGGGAAAGTTTATTTGATATGAGACTATGAAA 4648
Qy 2001 AATATATTAAATCAAGATAGTATTATTTGCCAGGAGCCCGCTACTTTTATGAAAT 2060
Db 4649 GATGATGAATGAAGCAGTACGATTTATCTCATGCGGACCACTTACCTATATGCAAG 4708
Qy 2061 CATTATCAAGGAAAGAAACAAATATTTGTTCTAGCAAAAGAAAGTATGTTGACATG 2120
Db 4709 TATTCAACTAGGTAAATATCCGATAGTTGTTCCACGGCAATGAAATTTGATGAGCATA 4768
Qy 2121 TAAATGATCATCAA-GTAGAGTTTGTAAAGAAATTTTCAAGATAATAATTTTATTT 2179
Db 4769 TAAATGATCATCACTTTGGTAAAGTAAACAGTTTGTAAAGAGGATCTCATTTGATTT 4828
Qy 2180 --ATAGAAATATAGATGATTTGTTTCAAAATATTTAGTTCCTTAAAGAACTAACT 2237
Db 4829 TGTGGAAGATGTTGAAGACATCTCGAAATATTTATTTAGTTCCAAATTTTCAGATACCT 4888
Qy 2238 TTACATCAATATAATTTTGTGTAAGATTAACAAATAGTTTGAAGAAATTTAATG 2297
Db 4889 TACAAAAAATGTAAATCAACAACTGAATTCATAAAATTTATTCAGTGTGAAATTTACC 4948
Qy 2298 AG 2299
Db 4949 AG 4950

RESULT 23

US-09-815-242-9587
; Sequence 9587, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815.242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9587
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1368)
US-09-815-242-9587

Query Match 6.1%; Score 423.2; DB 9; Length 1368;
Best Local Similarity 58.6%; Pred. No. 4.1e-50;
Matches 794; Conservative 0; Mismatches 548; Indels 12; Gaps 3;

Qy 14 ATTGGCAATTATTGATATGATGACAGTTGCAATTTCTGCAATCTTAAAGATCATATACC 73
Db 27 ATTGGCAATAATTCAGAGTTTCTTGTATTTTATTTAGCTTATCTACTTAGTCTGTGAG 86
Qy 74 AATGCTGATTAAATCGTTCTGGAATTTTATCATAAATGATGTTTCATTTATTTTGCATT 133
Db 87 AGAAGCGAGATTGTTTCAACACACTATTTGCACTTTATATCTCTCATTTATTTTGTCTT 146
Qy 134 TTTTATATCTCGTATGCCAGTTTGAATTTGAGTATAGAGGTAATCTGATAGAGTTTGAATA 193
Db 147 TTATATCAGTGATTATGAGCAGGATTTCTTAAAGAGAGATATTTGATGAACCTTGCCA 206
Qy 194 AACATTTAATATAGTATATATTTTTCGAATTTTTCGCGAGTATCATTTTGTGGA 253
Db 207 GACATTTGAATAATATCTTATTTCTTGCACTAGCGATTGGTATTTCTAAATTTTCTTAGA 266
Qy 254 GAATAATTTGCACTTTTCAAGAGCTGTCGCGTATTTTACATTAATTAATTAACCTCGTTT 313
Db 267 GGATCGATTAGTATTTCCAGACGAGCATGATTTACTTCTCCTCACATTAATGCTCTCTT 326
Qy 314 GGTATACCTATTATTAAGCAGTTTAAAGGATGAGTCTTCTATTTTCGACAAT 373
Db 327 AGTCTATGCTTAAACCTATTTATCAAGTGTATTTGAGCGGCTTATCCCACTTTAA 386
Qy 374 CTATCAAAAAAGACGATTTCTAATTAACCGCTGAAACGATGGAAGAAATATGCAAGTTT 433
Db 387 AGGAAGTAAGAGATTTCTTACTTACAGCAACTTCTCGTGTGCAAAAGGTACTGGATAG 446
Qy 434 ATTTGATACATAAACAATTTCAAAAATCTTTGTCATTTGTTAGTATGATGATCAGA 493
Db 447 ATTAATAGAAATCAATGAGGTTGTTGGAGGTTGGTAGCGCTCAGTGTCTTAGATAACC 506
Qy 494 AATAGATAAAATTAATTTTATCATTCGCTCTATTTATTTCTGTGGAAGAGCTATAGAGTT 553
Db 507 AGATTTTCAGCATGATTTTAAAGGTAGT-----AGCAGAGGGGGAGATAGTAAACTT 560
Qy 554 TTCAACAGAGGAAGTGGTGCACCACTTTTATAAAATCTTACCAAGTGAAGTTTTTAGACGT 613
Db 561 TCGGACTCATGAGGTGGTGCATGAAGTCTTTTATCAATCTTCCAGGTGAAAAATACATAT 620
Qy 614 AAGCAATTTGTTTCAGATTTTTCAGTTTGTAGGTTTGTAGTATGATGAGGTTGATATTAATC 673
Db 621 TGGAGAGCTTGTCTCTCAGTTTGAACGATGGGAAATGATGATATAGTCAATCTTAAATGC 680
Qy 674 ATTCGGTTTTTACTGCGTTGAAAAACAAAAATCTCAACTCTAGGTGACCATAGCATTTGT 733
Db 681 TTTTGTATCGTAGTTTGGCAGCTAACCAAGCAAAATTCGTGAGATGCGCAGGGTTAAACGTTGT 740

| | | | |
|----|------|------------------------------------------------------------------|------|
| Qy | 734 | AACTTTTTCACAAATTTTATAAGCCTAGTCATATCANGAAGAAACGACTTTTGGATAT | 793 |
| Db | 741 | GACTTTTTTCTACAAATTTTATAAGACTAGTCATGTAAATTGCTAAGCGGATTAATGATAT | 800 |
| Qy | 794 | ACTCGAGCGGTAGTCGGGTAAATATTTTGGTGATAGTTTCTATTTTTGTAGTTCCAAAT | 853 |
| Db | 801 | CGTGGGTGCATTTGTTAGGGCTGACTATATGTGGTTTGTAGTCAGTATTGTACTGGTTCCTTT | 860 |
| Qy | 854 | TATTCGTPAGAGATGGTGGACCGGCTATTTTGTCTCAGAAACGAGTTGGACAGAAATGGAACG | 913 |
| Db | 861 | GAATTCGAAAGGATGGGGGATCTGCTATTTTGTCTCAGACGCGTATAGCAAAAATGCTCG | 920 |
| Qy | 914 | CATATTTTACATTTCTACAAGTTTCGATCGATGCTATGTTGATGCTCAGGAGCGCAAAAAGA | 973 |
| Db | 921 | TCAGTTTCACTTTTATTAAGTTTTCGCTCTATGTGTGTAGATGCCGAGCGCAAAAAGAGA | 980 |
| Qy | 974 | CTTGCTCAGCCAAACACAGATGCAAGGGTGGGTATGTTTAAAAATGGAAAAACGATCCT | 1033 |
| Db | 981 | ACTCATGGAACAAAATACCATCGAAGGTGGAAATGTTT-----AAGGTGGACGATGATCCT | 1035 |
| Qy | 1034 | AGAATTACTCGAATTGGACATTTTCATACGCAAAAAACAAAGTTTGTAGACAGATTACACAGTT | 1093 |
| Db | 1036 | CGTATCACGAAAATTTGGTTGTTTATATACG-GAAGACTAGCTTGCACGAGCTACCCACAGTT | 1094 |
| Qy | 1094 | TTATAATGTTTTTAATTCGCGATATGACTAGTTTGGTACGTCACCTACCTACAGTTGATGA | 1153 |
| Db | 1095 | TTATAATGTTCTAAAGGAGATATGAGTTTGGTTGGCACGCGCCACCAACAGTGGACGA | 1154 |
| Qy | 1154 | ATTTTGAAAAATATATCTCTGGTCAAAAGACGATTCAGTTTAAAAACGAGGATTTACAGG | 1213 |
| Db | 1155 | GTATGACACTATATACCCAGAACAAAACGCCGACTAAGTTTTTAAACCTTGCATAAACAGG | 1214 |
| Qy | 1214 | TCTCTGCAGGTAGTCGGTGGTAGTAATATCACAGCTTCGACGACGATGTTTGGTGGGA | 1273 |
| Db | 1215 | TTTATGCGCGGTTCAGCGGCGAAGTGGAGTCAAGAAATTCGATGAAGTTGTCTCAAAATTAGA | 1274 |
| Qy | 1274 | CTTAGCATACATTTGATAAATTTGGACCTATCTGGTCAGATATTAATAATTTTATAAGACAGT | 1333 |
| Db | 1275 | TGTGGCCCTATATTAATGTTGGACAACTCTGGAAAGATATTAATAATTTTATTGAACAGT | 1334 |
| Qy | 1334 | GAAGTTGTTATTTGTGAGAGAGGGAAGTAAGTAA | 1367 |
| Db | 1335 | TAAAGTAGTATTTATGAGATGATGAGCGCAAGTAA | 1368 |

RESULT 24

```

RESU01 24
US-10-804-408-178
; Sequence 178, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 178
; LENGTH: 2581
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
US-10-804-408-178

```

Query Match 6.0%; Score 419.8; DB 18; Length 2581;
Best Local Similarity 51.9%; Pred. No. 1.6e-49;
Matches 1100; Conservative 0; Mismatches 972; Indels 47; Gaps 5;
QY 3063 GAATATTTTATTTTAAAGTACTTCTGGTGTGATATTTTATTCAGAGCAAAAGTATGTA 3122

| | | | |
|------|----|-------------------------------------------------------------------|------|
| 1569 | Db | GTATTAATTATCTCAACTCTTTGGTAT-----AGTGGGAAGGG | 1601 |
| 4188 | Qy | ACAAAAATGAATGATTTAAATTTTCAGGTATTGTACCAATTTATAATGTCACAAATATCT | 4247 |
| 1606 | Db | CTAAAAAATGAAGAAAAAGTAAACAGTCATTATACCTTATATACAACCTCAGAAGCATACCT | 1665 |
| 4248 | Qy | TGATAAATGTTATTAACAGTATTATTAAACCAACATATACTAAATTTAGAGGTTATTTCTCGT | 4307 |
| 1666 | Db | TAAAGAATGTGTGCAATCCGCTACTACACACAGACTCATCCATTGATGAAGGTTATACTTAAT | 1725 |
| 4308 | Qy | AAATGATGGAAGTACTGATGATTCTGAGAAAAATTTGCTTAAACTATATATGAAGAACGATGG | 4367 |
| 1726 | Db | TGATGATGGATCCACTGATTAATAGTGAGAAATTTGTGATTAATTTATCTCAGAGAGATAA | 1785 |
| 4368 | Qy | AAGAATTTAAATAATTACAAAGAAAAATTAAATGCGGCTTAGCAGATGCTCGAAAATTTCCGACT | 4427 |
| 1786 | Db | TCGCATACTTTGTTATTTTCATAAAAAATGAGGGGTCTCTTCGCAAGGAACCTAGGTCT | 1845 |
| 4428 | Qy | AGAACATGCACACAGGTAATATATTGCTTTTGTTCGATTCTTGATGACTATATATAGAAGTTGC | 4487 |
| 1846 | Db | AGATAAATCCACAGGAGAAATTCATAACATTTGTGGATAGTGATGATTTTGTAGCACCGAA | 1905 |
| 4488 | Qy | AATGTTCCGAGAGATGCATGATTAATAATAAATCTGAGTATAAATCTGCCGATATACGAGATAGA | 4547 |
| 1906 | Db | TATGATTTGAAATTAATGTTAAAAAATTTAACTCACTGAGAAATGCTGATATACGAGAGTAGA | 1965 |
| 4548 | Qy | TTTTTTGTTAGTAGACGAAACCGGTTATACAAAGAAAAAAGAAATAGTAAATTTTTCATGT | 4607 |
| 1966 | Db | TTTTTGATTTTCGAATCAGAGAGATTTATGAAGAAGAAAAAGACGAACTTTTATATAAGT | 2025 |
| 4608 | Qy | CTTAACCAGAGAGACAGACTGTAAAGAGAAATTTTGTGTGAGGATCTTAATATACAAATAATGT | 4667 |
| 2026 | Db | TTTTTAAGAATAATTAATCTTTTGAAGAATTTTATCAGGTAATAGAGTGGAAATATTTGT | 2085 |
| 4668 | Qy | TTGGTGCAGGCTTTATTCACGAGATATTATAAAGATATATAAAATTCCAAATTTAATAATAG | 4727 |
| 2086 | Db | TTGTACAAAAATTATATAAAAAAAGTATAAATGGTAACTTCGAGGTTTGA---TCGAACTT | 2142 |
| 4728 | Qy | AAGTATTTGGTGAGGATTTGCTTTTAAATTTGGAGGCTCTCGAACATGTATAACGTTAGT | 4787 |
| 2143 | Db | AAAAATTTGGTGAGGATTTACTTTTAAATTCGAACTCTTATGTCAAGAGACCGGTATAGT | 2202 |
| 4788 | Qy | AGTTGATCTAGAGAAATTAATTATAATATTATGTCATTCTGTAACAGTTTCGGCTTAATTAATCA | 4847 |
| 2203 | Db | CGTAGATACGACTCTCTTCTTATATCTTATCGAATTTGTAAAACTTCTGTATGTAATCA | 2262 |
| 4848 | Qy | GAAATTTCTTAATAAATAATTGATTTAGTCAACAGATTCGAGAAATPACCCCTTTTAAGTT | 4907 |
| 2263 | Db | GAAATTTCAACGAAAACTCATTTAGATTTTATAACAAATTTTAAATGAATAAGTAGTTTGGT | 2322 |
| 4908 | Qy | AAAAAGAGAGTTTAGTCATTTATTTTGATGCAAAAGTTTATTAAGAGAGAGGTTAAATGTTTT | 4967 |
| 2323 | Db | TCCTGCCAGATAGCTAAATTTATGTTGAAGCGAAATTTTAAAGAGAAAGAATAAAGTGCTCT | 2382 |
| 4968 | Qy | AAACAAAAATGTAATTCACAGATTG---TTTGGATAATGAGTTCTTCGCAATATATTAGAGTC | 5024 |
| 2383 | Db | CCGAAAAATGTTTGAAATTTAGGTAGTAATAATTGACAAATAAAATCAAAGTACAACGAGAGAT | 2442 |
| 5025 | Qy | TTATCGAAAAAGAAATACGTAGATATCATTTTATTAAGCGAAAAAGATATTTTATCAAGAAA | 5084 |
| 2443 | Db | TTTTTTTCAAAAGACATAAATCATACCCGGTTCTATAAAAGCGGTCAAAATCTATCATTTAAA | 2502 |
| 5085 | Qy | GCATTTAGTTAGGTTGTAATTTGATGAAATTTTTCGCCCTAAACTATCTATGTAATGTTATATAA | 5144 |
| 2503 | Db | GGGATTTATTAAGCTTTTATTTAATGAATGTTTCACTTAACATATATGTTATGGCATATAG | 2562 |
| 5145 | Qy | GAAATTTCAAAAGCAGTAG | 5163 |
| 2563 | Db | AAGATTTTCAAAAAACAGTAG | 2581 |

```

; Sequence 179, Application US/10804408
; Publication No. US20040253617A1
; GENERAL INFORMATION:
; APPLICANT: FANRONG, KONG
; APPLICANT: GILBERT, GWENDOLYN
; TITLE OF INVENTION: MOLECULAR TYPING OF GROUP B STREPTOCOCCI
; FILE REFERENCE: 675002-2001
; CURRENT APPLICATION NUMBER: US/10/804,408
; CURRENT FILING DATE: 2004-03-19
; PRIOR APPLICATION NUMBER: PCT/AU02/01281
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: AU PR 7749
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 179
; LENGTH: 2577
; TYPE: DNA
; ORGANISM: Streptococcus agalactiae
; US-10-804-408-179

Query Match          5.5%; Score 386; DB 18; Length 2577;
Best Local Similarity 51.6%; Pred. No. 8,9e+45;
Matches 1092; Conservative 0; Mismatches 975; Indels 51; Gaps 7;

Qy 3063 GAATATATTTTAAAGTACTTCGTGGTGAATTAATTTTATCCAGAGCAAAAGTAGTA 3122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 492 GATTATCTTTATTTATGCTCTTGGGTACTTTATTTTAGTACCAACCAATGGTATCAG 551

Qy 3123 TTTTATTAATTTTATGAATTAATTTTATTTTCATATAAAATTTTGGAAACTAAGCTA 3182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 552 TTTTAAATATTACCATTATAGTCTTATTATTTCTTGGGAAGAGTGAGTTAGAAATATCT 611

Qy 3183 ATATTAATAAAATGAAATTTTATTTGTTTTTATTATGCTGTATATTTATGTTTGTTCAGTA 3242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 612 ATAAGCAATCTTCAATACTATTTCTGCTTGGTATTTATTTTATTTTGGCAATCTC 671

Qy 3243 GTCACAAGTAGTTTGTGAAATAAAATTTTGAAGAATTAATTGCAGATTTTACTGCTCCC 3302
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 672 ATTAGAGGTACTCAAGAGGATATAAGCTTTTCAGCGATTTATTGCTGAGCTATTAAAACTA 731

Qy 3303 ATAATTTGGATTATTCGAATAAGTATTAATTTCTGATTCATTTATAATATTTGATTAT 3362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 732 ATTAGTACAGGATATGCTTTATTTTATTAATTAATTATAGAAAGCTGATTTTAATAGT 791

Qy 3363 AAAAAATTAATAAATAGTATCTTTTTTAGTTTTTTTAGTTTTTATTAGTATATCTGCATTG 3422
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 792 TCAGTTTGAAGGAATGTGGTAAAGGTTAACTATTTTGTGTTGTTCTTTATAACAGTTTAA 851

Qy 3423 TATATTTTCAAATGGGAAGATATTTGATTTTTTATAGACAGACACCTTTATAGACTAGAC 3482
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 852 TATTTATTTTTTCCAAATGAAATTTACTACATTCCTAGGAGAGATTTATTTTCCAATGAA 911

Qy 3483 TATCTTATAACAGCGCTCAAAACAAGGTGGTTGGCTTTATGAACTATCTAGCTTAAAT 3542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 912 T----GAAATCCTCTATGAAGGTAGACTTACTGCAATTTTGGAGTATGCAACACTATTA 968

Qy 3543 ACCACTACAATTATAGTTTCAAATTCGGTTAAATCTTTGCACTTATAAAAAATAAAATGCAA 3602
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 969 GGTGAGTTTATTTTATTTCACTTATCCGATATATTTTAAAAACAGCAGAGGTATGGAGAA 1028

Qy 3603 CAATTTTTTTTTCTGTGCTGTGCTTTTATACCGATCTATTTAAGTGGATCGAAGATTGGT 3662
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1029 AATATTTTTTATCACATAATTCCTAGTTTGTGTCATATTTTGAAGGGCAAGAAATTTTC 1088

Qy 3663 AGTTTATCGCTAGCAATATTAATTTATATGCTCTGTATGGAGATATATAGTGTGAAATTTT 3722
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1089 CTAATTTGTATGATTAATTTTATTAGTTATTTACTCTTAGAATAATCATTAATTAATTT 1148

Qy 3723 GCTTGGATAAAAAAGCTAATA----GTAATTTTGTAACTACTTATTTATTTAAATPACT 3779
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1149 AACCTAAAAAATTACTAAAAAAGCTGCTTTTTCGATAATTTATAGGGATAATTAATTTATTG 1208

```


This Page Blank (uspto)


```
Qy 1911 CCGACGAAATATTTATTCACACAGGATATTCGTGACTATATTCAGAAATATTCGACGATATA 1970
Db 1825 ATAAAAAAATNTAATAAAAAATATAATATATATATAATATAAAAAATATAAAAAATAAAAA 1766
Qy 1971 AAAAAATTCAGTTACAAAGAAATGCAACAATATAATTAACAATCAGAAAGTAGTTATTTT 2030
Db 1765 ATTAAATATATAATATANAATTTAAATATAATATAATATAATATAATATAATATAATATAAT 1706
Qy 2031 GCCACGAGGCCCGCTACTTTTATGAATTCATPATCCAAAGGAAAAAACAATTAATGT 2090
Db 1705 ATTATATAAATAAAAAATNTAAATTTATATANAATAATAATTAATAAAAAATAAATAA 1646
Qy 2091 TTCTTACACAAAAAGTAGTGGCAACATGTAATGATCATCAAGTAGAGTTTGTAGAA 2150
Db 1645 TAATTANNAAAAAAATAATTTATTTTNAATTTATATATAATAATAATTAATTTTAAATA 1586
Qy 2151 GAATTTTACAGATAAATAATTTTATTTATAG-----AAAAATAGATGATTTC 2200
Db 1585 TNATTATAAATAATATATATATTTTTTATTTTTTATTTTTTATTTTTTATATAATATATTT 1526
Qy 2201 TTGAAAAAATATGGAAGTTTCTPAAGCAAACTAACTTTTACATCAATAATAATTTTTTT 2260
Db 1525 TAATAATTTTATAAAAAATAATATAANATAAAATATATTTTTTTTTTTTTTTTTTTT 1466
Qy 2261 TGTCAAGAGATTAACAACAAATAGTTGAAAAATTTAATGAGGATCAAGAAATGAATAATA 2320
Db 1465 TTTAAAAATATATATTTATATAAATAAAATTTANAAATAAAAAAATAATAAATAATTTT 1406
Qy 2321 AAAAGATGCAATATTTGATATGCTTATCATATAATTTTCTCAGATTTTCTGAGAGGGA 2380
Db 1405 ATTTTATTAATATTTAAANATAATTTAAATTTTNNANATAATAATAATATAATAAATAT 1346
Qy 2381 TACAGATATATCATCTCTCTCAGGAGAAATGCACACCATAGTTCTCTCAGAAATCCTG 2440
Db 1345 NATAAATTAATAATTAATAATTTATTTTAAATTTTAAATTTTAAATTTTAAATAATA 1286
Qy 2441 TATAATTTATTTAAATATCTCAGGATTTATATGTTGAAATTTACAAAAGTAGCAAAAA 2500
Db 1285 AAAAAAATAAATTTATATAAATAATAATAATAAATAAATAAATAAATAAATAAATAA 1226
Qy 2501 TATAAGAAATAGGATATATGACGAGTTAAAGTTTACAGATTTATTTCTTAATATATCA 2560
Db 1225 TATATAATTAATATATATAAATAAATAAATAATTTAAATAAATAATTTAATAAATAT 1166
Qy 2561 GAAAAAATAATGATAATGTAAGTCTTTTGAAGATTTTAAAGAAATGTAACGAGCTTTTGAA 2620
Db 1165 AATAATATAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1106
Qy 2621 TACTATTTACAAAGATTTGTTGTTTATGATAGAAATAAAAAACATGGCTTAAGAAATAGAT 2680
Db 1105 AAATNTATTTTATTAATTTAAATAATAATAATTTATTTTATATAATAATTAATTA 1046
Qy 2681 TTGGTTCTAATGGGTTTCGCTCCACATGATTTTGTGGCAATTTCTTTTATCAAAATGAA 2740
Db 1045 TATTTNTTATATATTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 986
Qy 2741 ACGAAACAGCTTATTTTATTAAGTAATCTAAATCTCCAGATGAACACTTTTATACAGCAA 2800
Db 985 --TANATATTTTATTAATTTATTAATAATAATAAATAAATAAATAAATAAATAAATAA 928
Qy 2801 TTATAGAAAAATATGAATTTTCAATAAATATCTAAATATGGAATTTTAAAGATATATA 2859
Db 927 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 868
Qy 2860 AAGTGGAAAAAATCAACATCTTCTCTTATGCTTTTACAGATGATCTTATTTGATGATTC 2919
Db 867 AAAAAATANAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 808
Qy 2920 CTAATATCAAGAAATTTAGGTTTTTTTATTTGCTAGAAAGTTAAAAATAGAAAAATAATCT 2979
Db 807 TAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 748
```

```
Qy 2980 ABAATTAAGAAATATTTACTAAAAAATAAATAGTTGATTTTGTGAGAGTAATGTATGT 3039
Db 747 ABAATTAAGAAATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 688
Qy 3040 TTAATAATTTTAAAA-----TATGACCCGGAATATTTTATTTTAAAGTACTCTCTGGTTGAT 3094
Db 687 TTTTATTTATTTATTTATTTATTAATAATAATAATAATAATAATAATAATAATAATAATAA 628
Qy 3095 TATTTTATTTCCAGACAAAGATATGTAATTTTATTAATAATTTTATGAAATTTTATTTAT 3154
Db 627 AATTATNAATTTATTAATAATATATATAAATAAATAAATAAATAAATAAATAAATAAATAA 568
Qy 3155 TCAAT--ATAAATTTTGAATACTAAGCTAATAATTAATAAATAAATAAATAAATAAATAA 3212
Db 567 TNATAAATAAATAATATATATATANTTAATAATAATAATAATAATAAATAAATAAATAA 508
Qy 3213 TTATGCTCTATATTTATGTTTGTTCAGTAGTCACAAGTAGTTGTTGTTGAAAAATAATTTT 3272
Db 507 TTATTAATAATTTATTTATANAATTAATAATTTATTTATTTATTAATAATAATTTTATTA 448
Qy 3273 GAAAGATATTTGACAGATTTTACTGCTCCCATTAATTTGGATTTTGCATAAATGATATAT 3332
Db 447 TTTATAAATAATAATAATAATAATAATTTAAATTTAATAAATAAATAAATAAATAAATAA 368
Qy 3333 AATTGCTATTCATTATAAATAATTTGATTTATAAATAAATAAATAAATAAATAAATAAATA 3392
Db 387 TTATTAATAATTTATTAATAATANNNTAATAATTTNTTTTATAAATAAATAAATAATA 328
Qy 3393 TTTTATGTTTATTTAGGTATATCTGCATTTGTAATTTATTTTCAAAATCGGAAAGATATTGA 3452
Db 327 TAATAATAATAATAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 268
Qy 3453 TTTTATGACAGACACCTTAT 3472
Db 267 TAATTAATAATAAATAATAT 248

RESULT 2
CG753083
LOCUS
DEFINITION
1896 bp DNA linear GSS 24-OCT-2003
P048-1-C01.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.
CG753083
VERSION
CG753083.1 GI:37977199
SOURCE
GSS.
ORGANISM
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE
1 (bases 1 to 1896)
Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
An integrated physical and genetic map of the nematode Pristionchus
pacificus
Mol Genet. Genomics 269 (5), 715-722 (2003)
22835951
12884007
COMMENT
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
Location/Qualifiers
1..1896
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of
```


the genomic DNA with EcoRI and cloning into the BAC vector."

ORIGIN

```

Query Match          2.4%; Score 167.4; DB 9; Length 1896;
Best Local Similarity 46.8%; Pred. No. 1.1e-15;
Matches 761; Conservative 0; Mismatches 830; Indels 35; Gaps 9;

Qy 2050 TTTTATGAATTCATTATCCAAAGGAAACAAATTAATTTGTTCTAGACAAAAGTA 2109
Db 219  TNATATTTATTTATTAATAAATAAATAATATTTATATTTATTAATTAATAAANT 278
Qy 2110 TGGTGAACATGTAATGATCATCAAGTAGAGTTTGTGAAGAAGATTTTACAGATAATA 2169
Db 279  TANATATTANANATATTTANNANAAATTTATTTATTTATTTATTTATTAATAATTAA 338
Qy 2170 TATTTATTTATAGAAATATAGATGATTTGTTTGAAGAAAT-TAATGAAAGTTTCTAAGC 2228
Db 339  TATTTATTTTATAAANAATAATTTATANNATATATTAATAATAATAAATTTATTTAA 398
Qy 2229 AAACCTAACTTTACATCAAAATAAATTTTCTGGAAGATTAAAAACAATAGTTGAAA 2288
Db 399  TTATAAATNATTTTAATAATTAATAATATATATTTATTTATTAATAAATAATA 458
Qy 2289 AATTAAATGAGATCAAGAAATGAATAATAAAGATGATATTTGATAAGCTTAT 2348
Db 459  AAAATATTTAATAATAATAAATTTATATNTATATAAATAATTTAAATAAATTTATAT 518
Qy 2349 CATTAATTTTCTCAGATTTTACTGGAGGGATACAGATATTTATCATCTTCTCAGAG 2408
Db 519  AATTTTATATATTTATATATAATAATAATAATAATAATAATAATAATAATAATAATA 578
Qy 2409 AATGCACACCAATTAATGTTTCTTCAAGATACCTGTTAATAATTTTAAATATTTCTCAGGAT 2468
Db 579  ATTATATAAANNNATTTATTTATATAATAATTTTAAATNATNATTTTATTAATTAAT 638
Qy 2469 TATATGTTGAATTTACAAAAGATGAGCAAAATAATNAAGAAATPAGGATATGAAAGAG 2528
Db 639  ATATTTNNATATTTATATATATTTATTAATAATAATAATAAATAAATAATAATAAT 698
Qy 2529 TTAATGTTTACAGATTTATTTCTTCAATATATACAGAAATACTAATGA--TAATGTAAGTT 2586
Db 699  ATATTTTAAATTTTATTTTAAATTTTATTTATTAATAAATTTTAAATTTATTTATTA 758
Qy 2587 TAGAATTTTATTAAGAATGATCCAGCTTTTGAATACTATTTTACAAAGATTTGTTTAT 2646
Db 759  TAATTTAATAATATTTAAATTAATAAATAATTTATTTATTTTAAATTTTATTTTAA 818
Qy 2647 TGATAGAATAAAAAACATGGTCTAAGAATAAGATTTGGTCTTAATGGGTTTCCGTTCCA 2706
Db 819  TNATAAATATATATAAATAATAATAATTTATTTATTTATTTATTTATTTTATTTTAA 878
Qy 2707 CATGATTTTGGCAATCTTTTATCAAAAGAA--AACGAACAGCTTATTTATTAAGTA 2765
Db 879  ATTTATTTTATTTATTAATAATAATTTTAAATAATTTTAAATTTTATTTTAAATTAATA 938
Qy 2766 ATCTAATGTCAGATG-----RACATTTTATACAGACAATTTAGAAAATATGAAATTT 2820
Db 939  AATTTATTTTAAATATATTTTAAATAATTTTAAATAATAATAATAATAATAATAATA 998
Qy 2821 TCAATAGATTTATCTAAATATGGAATTTAAGATATATAAAGTGG----- 2865
Db 999  AAAATTTTATTAATAATTTTATTTATATAATAATAATAATAATAATAATAATAATAAT 1058
Qy 2866 ---AAAAAATCAACATCTTCTTATGCTTTTACAGATGATTTATTTGATGAATTTGCTA 2922
Db 1059 ATAAATAAATAAATTTATTTAATAATTTAAATAATAATAATAATAATAATAATAAT 1118
Qy 2923 AATCAAGAATTTAGGTTT---TTTATTTCTAGAAAGTTTAAATAAGAAATAAATCT 2979
Db 1119 TATAATAATTTATTTTATTTATTTATTTTATTTATTTATTTATTTATTTATTTA 1178
Qy 2980 AAATTTAAAGAAATTTATTTACTAAAAATAAATAAGTTGATTT---TGTGAGAGTAAGTA 3036

```

```

Db 1179 AATTTTATTAATAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1238
Qy 3037 TGTTTAAATTTATTAATAATGACCGGAATATTTTATTTTAAAGTACTTCTGGTTGATTA 3096
Db 1239 TATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1298
Qy 3097 TTTTATTTCCAGAGCAAAAGTATGTTATTTTATTTTAAATTTTATGAAATTTAAATTTAT 3156
Db 1299 TTAATTTAAATAATTTAAAAAATAAATTTAAATTTATTTTAAATTTATTTATTTATTTAT 1358
Qy 3157 ATATAAATTTTTCAGAACTTAAGCTAATATTAATAAATGAAATTTTATTTGTTTATTTAT 3216
Db 1359 TATTTATATTNNNTAAATTTTAAATATATTTTAAATATTTTAAATAAATAAATAATTTAT 1418
Qy 3217 GGTCTATATTATGTTTGTGTTTCAGTAGTCACAGATGTTTGTGGAATAAATAATTTGAAA 3276
Db 1419 AATTTTATTTT-TATTTNATATTTTATTTATTAATAATAATAATAATAAATAAATA 1477
Qy 3277 GATTTATTCAGATTTTACTGCTCCCATATTTTGGATTTATGCAATTAATGTTATTTAT 3336
Db 1478 AAAAAAATAAATTTTATTTATTTATTTATTTTAAATAAATAATTTAAAAATTTATAT 1537
Qy 3337 TGTATTCATTTATTAATAATTTGATTTAATAAATAAATAAATAAATAAATAAATAAATA 3396
Db 1538 ATATTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1597
Qy 3397 TAGTTTATTTAGGTATATCTGCAATTTGTTATTTTCAAAATGCGAAAGATTTCTATTTT 3456
Db 1598 TTATTTATTTATTTATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1656
Qy 3457 TAGACAGACACCTTTATAGGACTAGACTATCTTATAACAGGCGTCAAAACAGGTTGGTTG 3516
Db 1657 TTTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1716
Qy 3517 GCTTTATGAACACTTCTACGTTAAATACCAGTCAATTTATGTTTCAATCCGTTAATCT 3576
Db 1717 TAATAATAATTTATTTATTTTAAATTTTATTTATTTATTTATTTATTTATTTATTTT 1776
Qy 3577 TTGCACCTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3636
Db 1777 TAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1836
Qy 3637 TCTATT 3642
Db 1837 TTTATT 1842

```

RESULT 3

```

CG750135/c
LOCUS      CG750135      1592 bp      DNA      linear      GSS 24-OCT-2003
DEFINITION P044-3-D09.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION  CG750135
VERSION     CG750135.1  GI:37971278
KEYWORDS   GSS.
SOURCE     Pristionchus pacificus
            Pristionchus pacificus
            Nematoda; Chromadorea; Diplogasterida;
            Rukaryota; Metazoa; Nematoda; Pristionchus.
REFERENCE  1 (bases 1 to 1592)
AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE   22835951
PUBMED    12884007
COMMENT   Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371

```


| | | | | | | | | | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|-------------------------------------------------------------------|------|--|--|--|--|--|--|
| /tissue type="mixture of kidney and spleen" /clone_lib="MSMg01 Mouse Male BAC Library" | | | | | | | | | |
| ORIGIN | | | | | | | | | |
| Query Match 2.2%; Score 153.6; DB 9; Length 1542; Best Local Similarity 46.7%; Pred. No. 1.4e-13; Matches 587; Conservative 0; Mismatches 664; Indels 7; Gaps 4; | | | | | | | | | |
| Qy | 2604 | TGATCAGCCTTTGCAATCTATTACAAAGATTGGTTGTTATTTGATGAGATAAAAAACA | 2663 | | | | | | |
| Db | 155 | TTTATANAATTTANTTAATTTTAAAAAATTTTAAATATATATATATTTTAAAAAATATT | 214 | | | | | | |
| Qy | 2664 | TGGTCAAGATAAGATTTCGTTCTAATTCGGTTTCGCTCCCATGATTTTGGGAAT | 2723 | | | | | | |
| Db | 215 | TAATTTAATTAATTAATTTTAAATATATATTTTAAATATATTTTAAATATTTTAAATATT | 274 | | | | | | |
| Qy | 2724 | TCT--TTTATCAAAATGAAACGAAACAGCTTATTTATTTAAAGTAACTAAATGCCAGAT | 2781 | | | | | | |
| Db | 275 | TTTAAATATATATATTTAAATTTTAAATATTTTAAATATTTTAAATATTTTAAATTTT | 334 | | | | | | |
| Qy | 2782 | GAACTATTATACAGACAAATTTATAGAAAAATAGAAATTTTCAAAATAGATTTCTAAAT-- | 2839 | | | | | | |
| Db | 335 | NTTAAATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAA | 394 | | | | | | |
| Qy | 2840 | ATGCAAAATTTAAGATATATAAAGTGGAAAAATCAACATCTTCTCCTATTGCTTTACAG | 2899 | | | | | | |
| Db | 395 | ATATAAAATTTTAAATATATATTTTAAATATTTTAAATATTTTAAATATTTTAAAT | 454 | | | | | | |
| Qy | 2900 | ATGATTTCTATTGATGAATTTGCTAAATGCAAGAAATTTAGGTTTTTTTAAATTTGCTAGAAAGT | 2959 | | | | | | |
| Db | 455 | ATTTTTTTTTATATATATAAATATAAATATATTTTAAATATTTTAAATATTTTAAAT | 514 | | | | | | |
| Qy | 2960 | TAAAAATAGAAAAATCTAAATTTTAAAGAAATTTATCTAAAAATATAAATAGTTGAT | 3019 | | | | | | |
| Db | 515 | TTTATATTTTAAATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAA | 574 | | | | | | |
| Qy | 3020 | TTTCTGAGAGTAAGTATTTTAAATTTTAAATATAGACCCGGAATTTTAAATTTTAA | 3079 | | | | | | |
| Db | 575 | TTTTTAAATATANATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTT | 634 | | | | | | |
| Qy | 3080 | GTACTTCTGGTTGATTTTATTTTCCAGACCAAAAGTATGCTATTTTAAATTTTAA | 3139 | | | | | | |
| Db | 635 | TTATTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAA | 694 | | | | | | |
| Qy | 3140 | GAAATTAATTTTATTCATATAAAAATTTTGAAGCACTAAGCTAATATAAAAAATGAAT | 3199 | | | | | | |
| Db | 695 | TTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT | 754 | | | | | | |
| Qy | 3200 | TTTTATTTTAAATAGGCTATATTTATGTTTGTTCAGTAGTCACAGTATGTTGT | 3259 | | | | | | |
| Db | 755 | TTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT | 814 | | | | | | |
| Qy | 3260 | TGAATAAAATTTTGAAGATTAT--TTGCAAGATTTTACTGCTCCCAAAATTTGATTTAT | 3317 | | | | | | |
| Db | 815 | TTAATTAATTTTAAATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA | 874 | | | | | | |
| Qy | 3318 | GCAATAATGATTAATAATTTGATTTCAATTTAATAATTTGATTTAATAAAAAAT | 3377 | | | | | | |
| Db | 875 | TATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA | 934 | | | | | | |
| Qy | 3378 | AGTATCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT | 3437 | | | | | | |
| Db | 935 | NTTTATTTTAAATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT | 994 | | | | | | |
| Qy | 3438 | GGGAAAGATATTTGATTTTAAAGACAGACCTTATAGGACTAGACTATCTTATAACAGGC | 3497 | | | | | | |
| Db | 995 | TTTTAAATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA | 1054 | | | | | | |
| Qy | 3498 | GTCAAAAACAGGTTGGTTCCTTATGAACATCTCCTACGTTTAAATACCACTACAAATATA | 3557 | | | | | | |
| Db | 1055 | TTATTTTAAATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT | 1114 | | | | | | |
| Qy | 3558 | GTTTCAATTCGTTAACTTTGCACTTTATAAAAAATAAATGCAACAATTTTTTTCTTG | 3617 | | | | | | |

| | | | | | | | | | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|----------------------------------------------------------------|------|--|--|--|--|--|--|
| Db | 1115 | TTTTTTTATATTAATATATTTTATTAATTTATATATTTTATTAATTTTATTTT | 1174 | | | | | | |
| Qy | 3618 | TGCTTTGCTTTTATACCGATCTATTTAAAGTGGATCGAGAAATTTGATGTTTATCGCTAGCA | 3677 | | | | | | |
| Db | 1175 | AATATNTAATATATATATATATTTATTTTAT-NATTATTTTTTTTTTTTNTTATTTA | 1233 | | | | | | |
| Qy | 3678 | ATATTAAATATATGCTTGTATCGAGATATATAGGTGGAAAAATTTGCTTCGATAAAAAAG | 3737 | | | | | | |
| Db | 1234 | TTAATAATTTTATATATATTTTAAATTTTATTTTATTTTATTTTATTTTAAATTTATATA | 1293 | | | | | | |
| Qy | 3738 | CTAATAGTAAATTTTCTGAATCTACTTATTTTAAATCTCAATTTGCTTTTACCATGAA | 3797 | | | | | | |
| Db | 1294 | TTTTTATATATATATATATATTTTAAATTTTAAATTTTATATATATTTATTTTAAATAT | 1353 | | | | | | |
| Qy | 3798 | ATTTTGGCTGTTTATATTTCTAGAGAAATCAAGTAACGAAGCTAGATTTTATTTATTC | 3855 | | | | | | |
| Db | 1354 | TTACTATNTTTTATTTTAAATTTTNTATNTATATATATATATTTTATTTATTC | 1411 | | | | | | |
| RESULT 8 | | | | | | | | | |
| CL509408 1758 bp DNA linear GSS 01-APR-2004 | | | | | | | | | |
| SAIL_811_H11.v3 SAIL Collection Arabidopsis thaliana genomic clone | | | | | | | | | |
| CL509408 genomic survey sequence. | | | | | | | | | |
| GSS. CL509408.1 GI:46006728 | | | | | | | | | |
| Arabidopsis thaliana (thale cress) | | | | | | | | | |
| ORGANISM | | | | | | | | | |
| Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; | | | | | | | | | |
| Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; | | | | | | | | | |
| rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids. | | | | | | | | | |
| REFERENCE 1 (bases 1 to 1758) | | | | | | | | | |
| AUTHORS Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D., Dietrich,B., Ho,P., Racwaden,J., Ko,C., Clarke,J.D., Cotton,D., Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B., Mitzei,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A. | | | | | | | | | |
| A high-throughput Arabidopsis reverse genetics system | | | | | | | | | |
| PLANT Cell 14 (12), 2985-2994 (2002) | | | | | | | | | |
| PUBMED 22356987 | | | | | | | | | |
| COMMENT Contact: Sessions A Applied Trait Genetics Syngenta Biotechnology Inc. 3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA Email: allen.sessions@syngenta.com | | | | | | | | | |
| ABRC Stock Number CS836276: T-DNA left border flanking sequences of Syngenta Arabidopsis Insertion Library (SAIL) lines are available through the Arabidopsis Biological Resource Center (ABRC). Sequences represent a pool of amplified genomic regions and not single contiguous sequences. | | | | | | | | | |
| Class: TDNA tagged. | | | | | | | | | |
| FEATURES | | | | | | | | | |
| source Location/Qualifiers | | | | | | | | | |
| 1..1758 | | | | | | | | | |
| /organism="Arabidopsis thaliana" | | | | | | | | | |
| /mol_type="genomic DNA" | | | | | | | | | |
| /ecotype="Columbia" | | | | | | | | | |
| /db_xref="taxon:3702" | | | | | | | | | |
| /clone_lib="SAIL_H11.v3" | | | | | | | | | |
| /notes="SAIL Collection" | | | | | | | | | |
| /note="T-DNA left border sequences were isolated using a modified TAIL-PCR strategy" | | | | | | | | | |
| ORIGIN | | | | | | | | | |
| Query Match 2.2%; Score 150.8; DB 9; Length 1758; | | | | | | | | | |
| Best Local Similarity 36.3%; Pred. No. 3.7e-13; | | | | | | | | | |
| Matches 635; Conservative 0; Mismatches 1084; Indels 31; Gaps 5; | | | | | | | | | |
| Qy | 1968 | ATAAAAATTTCTCAGTTACAAAGAAATGGAACAATATTAATTAACAATCAGAGTAGTTA | 2027 | | | | | | |
| Db | 27 | AAANAANAANTNTTAAATTTAAAAAATAAATAAATNTTATATANNAAAAAATAAAAAA | 86 | | | | | | |

QY 2028 TTGGCAGGAGGCCCGCTACTTTTATGAAATTCATATCCAAAGGAAAAAACAATTAT 2087
Db 87 AAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 146
QY 2088 TGTTCCTAGACAAAAAAGTAGTGTGAACATGTAAATGATCATCAAGTACAGTTTGTA 2147
Db 147 AAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 206
QY 2148 GAAGAAATTTTCAAGATAATAATATTTTATTTATAGAAAAATATAGATGATTTGTGAAA 2207
Db 207 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 266
QY 2208 AAAATTTTGAAGTTCTTCAAGCAAACTAATCTTACATCAAAATTAATATTTTTTTGTGAAA 2267
Db 267 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 326
QY 2268 GATTAATAAATAAGTTCAAAAAATTTAATGAGGATCAAGAAAAATGAATAAATAAAGAT 2327
Db 327 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 383
QY 2328 GCATATTTGATAATGGCTTATCATAAATTTTCTCAGATTTTCTGAGAGGGATACAGAT 2387
Db 384 ANAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 443
QY 2388 ATTATCATCTCTCAGAGAAATGCAACCAATAGTTCTTCAAGATACCTGTATATTT 2447
Db 444 AANNNNAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 503
QY 2448 ATTTTAAATATCTCAGATTTTATATGTTGAATTTTACAAGAGATCAGCAAAATATTAAG 2507
Db 504 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 563
QY 2508 AAAATAGGATATATGAACGAGTTAAATGTTACAGATTTATTTCTCAATATATACAGAAAAA 2567
Db 564 AANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 623
QY 2568 CTATGTGATAATGATCTGTTTGAATTTTAAAGATGATGCGAGCTTTTGAATATTT 2627
Db 624 AANNNNANAA-----NNAANAANAANAANAANAANAANAANAANAANAANAANAANA 675
QY 2628 TACAAGATGTTGTTTATTCATAGAAATAAACAATGCTTCAAGATAAGATTTGGTTC 2687
Db 676 NANNNAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 735
QY 2688 TAAATGGGTTGCTTCCACATGATTTTGTGGCAATCTTTTATCAAAATGAACGAAAC 2747
Db 736 NANNNAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 795
QY 2748 AGCTTATTTTAAAGTAATCTAAATGTCCAGATGAAT--ATTATACAGACAAATTATA 2805
Db 796 ANNNNNNNNAATAAANAANAANAANAATAAATAAATAAATAAATAAATAAATAAATA 855
QY 2806 GAAAAATAGATTTTCAAAATAGATATCTAAATATGGAATTTTAAAGATATATAAAGTGG 2865
Db 856 AATAATAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 915
QY 2866 AAAAAATCAATCTTCTCTTATGCTTTTACAGATGATCTTATTTGATGAATTCGTAAT 2925
Db 916 AATAATAATNAANAANAANTTATAAATAAATAAATAAATAAATAAATAAATAAATAA 975
QY 2926 GCAAGAAATTTAGGTTTTTTTATTTTGTGAGAAGTTTAAAAATAGAAAAATCTAAATTT 2985
Db 976 AATAAANTTAATATATATATATTAATAAATAAATAAATAAATAAATAAATAAATAA 1035
QY 2986 AAGAAAAATTTACTAAAAAATAAATAAGTTGATTTTGTGAGAGTAATGATGTTTAAAT 3045
Db 1036 AAAAATAATAAANAANTATATAAATAAATAAATAAATAAATAAATAAATAAATAA 1082
QY 3046 TATTTAAATATGACCCGGAATTTTATTTTAAAGTACTTCTGTTGATATTTTATTTTC 3105
Db 1083 TATTAATANAATAAANAANTNNNTTAAANTATAAANTNNNTTNTNTATAAATAAANTTTTA 1142
QY 3106 CAGAGCAAAAGTATGATTTTTTATTAATTTTATGAAATTTAATTTTATTTTCATATAAAT 3165

Db 1143 ATANATAATAAANAANAATAATANAANTNAATTTAANAATTTAATAAATNTTATTAATAAT 1202
QY 3166 TTTTGAAGAACTAAGCTAATAATTTAAAAAATGAAATTTTATTTGTTTATTTATTTGCTATAT 3225
Db 1203 TATTNAATAATATNTAANTATTAATAANTATTTATTAATAAATATTAATAAANAATAAT 1262
QY 3226 TATGTTTTTCTCAGTAGTACAAAGTATGTTTGTGGAATAAATAAATTTTGAAGATTAATTG 3285
Db 1263 TTTNTATTAANTATTAATTTAAAAATATAAANTNTTAANTTAANTTAANTTAANTTAANTTA 1322
QY 3286 CAGATTTTACTGCTCCCATTAATTTGGATTTATGCAATTAATGATTAATAATTTGTTATTCAT 3345
Db 1323 TAAAAATATNTNNNTATNTTATTTTAAATNTTAAATNTTAAATNTTAAATNTTATTTATTT 1382
QY 3346 TTATAAATAATTTGATTAATAAAAAATTAAAAAATAGTATCTTTTTTAGTTTTTTAGTTTTAT 3405
Db 1383 TNNNAANTATNTTATTTATTTAATTTTNTTATATTTATTAATAAATATATTAATTAATAA 1442
QY 3406 TAGGTATATCTGCAATGCTATATTTCAAAATGGGAAGATATTTGATTTTGAACAGAC 3465
Db 1443 AAAAAATAAATAAANAANAATAAATAAANAANAATAAATAAATAAATAAATAAATAA 1502
QY 3466 ACCTTATAGGACTAGACTATCTTATAACAGCGCTCAA-----AACAGGTTGGTTGGCTT 3520
Db 1503 AATAAATAAANAATATTTTATNTNANATTTATNTNANNTTANTNTNNTTNTTNTNNNT 1562
QY 3521 TATGAACATCTCTACGTTAAATACCACTACAAATTTATAGTTTCAATTCGGTTAACTTTGC 3580
Db 1563 TTATTTTATATTTTAAANNANTTAATNTAATNTAATNTAATNTAATNTAATNTAATNTAN 1622
QY 3581 ACTTATAAAAAATAAATGCAACAATTTTTTCTGCTGCTGCTTTTATACCGATCTA 3640
Db 1623 TTAATAAATAAATNANANTNTNTTTAATAATAAATAAATAAATAAATAAATAAATAA 1682
QY 3641 TTTAAGTGGATCGGAATTTGAGTTTATCGCTAGCAATATTAATTAATATATGTTTATG 3700
Db 1683 TATANTTTAAATNTAATNTAATNTAATNTAATNTAATNTAATNTAATNTAATNTAATNT 1742
QY 3701 GAGATATATA 3710
Db 1743 AAAAAATAA 1752

RESULT 9
CL509408/c

LOCUS SAIL_H11_H11.v3 SAIL Collection Arabidopsis thaliana genomic clone
DEFINITION SAIL_H11_H11.v3, genomic survey sequence.

ACCESSION CL509408
VERSION CL509408.1 GI:46006728
KEYWORDS GSS.

SOURCE
ORGANISM

Arabidopsis thaliana (thale cress)
Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE
AUTHORS

Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
A high-throughput Arabidopsis reverse genetics system
Plant Cell 14 (12), 2985-2994 (2002)

TITLE
JOURNAL
MEDLINE
PUBMED

22356987
12468722
Contact: Sessions A
Applied Trait Genetics

Syngenta Biotechnology Inc.
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA

Email: allen.sessions@syngenta.com
ABRC Stock Number CS836276; T-DNA left border flanking sequences of
Syngenta Arabidopsis Insertion Library (SAIL) lines are available

through the Arabidopsis Biological Resource Center (ABRC). Sequences represent a pool of amplified genomic regions and not single contiguous sequences.
Class: TDNA tagged.

FEATURES
source

```

Location/Qualifiers
1. 1758
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="SAIL_811.H11.v3"
/clone_lib="SAIL Collection"
/note="T-DNA left border sequences were isolated using a
modified TAIL-PCR strategy"

```

ORIGIN

| | | | | |
|---------------------------|--------|--------------------------------------------------------------------|--------|--------------|
| Query Match | 2.1%; | Score 149.6; | DB 9; | Length 1758; |
| Best Local Similarity | 36.2%; | Pred. No. 5.7e-13; | | |
| Matches 631; Conservative | 0; | Mismatches 1086; | Indels | 27; Gaps 5; |
| Qy | 2134 | AGTAGAGTCTGTGAAGAAGATTTTTCACAAGATAAATAATTATTTATTTATGAAATAATAGA | 2193 | |
| Dd | 1742 | ANTAATTAATTTTAATATAAAAATTTTTATNTATNATATATTNATPATNATTTTHANANTATA | 1683 | |
| Qy | 2194 | TGATTTTGTTGAAAAAATTAATGAAGTTTCTAAGCAAACTAACCTTTACATCAAATAATAA | 2253 | |
| Dd | 1682 | TANTTTAATATAATAATAANTANTTATATATTTATAAANAANANNNTNTATTTATTTTTTAA | 1623 | |
| Qy | 2254 | TTTTTTTGTGAAGAATTAACAAATAGTTGAAAAATTTTAAATGAGATCAAGAAATAGA | 2313 | |
| Dd | 1622 | NNTANTATTAATAAAATATATATANATTANTATATTTAANTNNTTTAAAAATATATAAAATAA | 1563 | |
| Qy | 2314 | ATAATAAAAAAGATGC-----ATATTTGATATAGCTTATCATAAATTTTTC | 2359 | |
| Dd | 1562 | ANNNAANAATANTANTANNNANTANATAATNTNANATAAANTAAATATNTTTATTTATT | 1503 | |
| Qy | 2360 | TCAGATTTTACTGGAGAGGGATACAGATATATATCATCTCTCCTCAGGAGAAATGCACACCA | 2419 | |
| Dd | 1502 | TTTTTAITTTATATATTTATATTTATTTTTTTTTTATATATTTNTTTTATTTATTTANTTTTTT | 1443 | |
| Qy | 2420 | TTAGTCTCCTTCAGAAATACCTGTATAATTTATTTAAATATCTCAGGATTTTATATGTTGAA | 2479 | |
| Dd | 1442 | TTTAAITTAATATATATTTATATAATAATAATANNNAATAAATNATPAANAAATANTTTNNA | 1383 | |
| Qy | 2480 | TTTTCAAAAGATGACGAAAAATATAAGAAA--ATAGGATATATAGAACGAGTTAAATCTTA | 2538 | |
| Dd | 1382 | AATAATATATATANTTTATNTTATANNATTTATANAATTTATANAATAAANATANNNATATATTTA | 1323 | |
| Qy | 2539 | CAGATTATTTCCTAATATATACGABAAAACTATTTGATAATGCTACTGTTTAGAATTTTTATT | 2598 | |
| Dd | 1322 | ATATATTTTTNNNNTAANTTTTANTTAANNTTATT---ATATTTTAATTAATANTTAATA | 1267 | |
| Qy | 2599 | AAGAATGTATCGAGCTTTTGAATACTATTTACAAAGATTGTTGTTTATTCATAGAAATAAA | 2658 | |
| Dd | 1266 | NANAAATNTNTANTTATAATTTATTAATAATANANTTAATAATANTANATAATTATTN | 1207 | |
| Qy | 2659 | AAACATGGCTPAGAANAAGATTTGGTCTAAATTTGGGTTTCGGCTCCACATGATTTTGTG | 2718 | |
| Dd | 1206 | AATAAATTTTTTAATFANNATTTAAATAATNTTAAATTTNANTNTNTATAFTATNTTTATTTAT | 1147 | |
| Qy | 2719 | GCAATTCUTTTTATCAAATGAAAACGACGCTTATTTTATTAAGTAATCTAAATGTTCCA | 2778 | |
| Dd | 1146 | NTATTAANAANTTATTTATANAANAATTTATANTTTAAANNATNTNATTTAANTTATT | 1087 | |
| Qy | 2779 | GATGAACATAATTTATACAGACAATTTATAGAAAAATATGAAATTTTCAAATAGATATCTAAA | 2838 | |
| Dd | 1086 | AATAAATTTATTTAATATATATTTATATATTAJANTTTTTTATATANTTTTTNTANNTAAA | 1027 | |
| Qy | 2839 | TATGGAATTTTAAGATATATAAAGTGGAAAAAATCAACATCTTCTCCTATGTCTTTACA | 2898 | |
| Dd | 1026 | TTTATATTTAATAAATNATANATAATTTNAAATAATATAATAATTAANTTTTATTTAAAAAAA | 967 | |
| Qy | 2899 | GATGATTCATPTGATGAATTCGTAATCGAAGAAATTTAGTGTTTTTTTATTTGCTAG---- | 2954 | |

Db 215 TATATTTTATATATAAATTTATATATATATAATTTATTTAAAAATATTTTATTAATAATAATTT 156

Qy 3371 AAAAAATAGTATCTTTTT 3388

Db 155 AAAAAATATATAATATAT 138

RESULT 12

AG350139 1489 bp DNA linear GSS 02-JUN-2004

LOCUS Mus musculus molossinus DNA, clone:MSMg01-146K14.TJ, genomic survey

DEFINITION

ACCESSION AG350139

VERSION AG350139.1 GI:47923449

KEYWORDS GSS.

SOURCE Mus musculus molossinus

ORGANISM Mus musculus molossinus

REFERENCE 1

AUTHORS Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.

TITLE BAC end Sequences of Library MSMg01

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1489)

AUTHORS Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.

TITLE Direct Submission

JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan

1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail: hattori@gsc.riken.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)

Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).

Tsukuba Institute, Bio Resource Center.

The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan

phone: 81-298-36-9189, fax: 81-298-36-9199

e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : TJ

LIBRARY

Vector : pBACe3.6

R.Site 1 : EcoRI.

R.Site 2 : EcoRI.

FEATURES

source

1..1489

/organism="Mus musculus molossinus"

/mol_type="genomic DNA"

/sub_species="molossinus"

/db_xref="taxon:57486"

/clone="MSMg01-146K14.TJ"

/sex="male"

/tissue_type="mixture of kidney and spleen"

/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 2.1%; Score 144.4; DB 9; Length 1489;

Best Local Similarity 46.7%; Pred. No. 3.6e-12;

Matches 643; Conservative 0; Mismatches 699; Indels 34; Gaps 5;

Qy 2115 AACATGTAAATGATCATCAAGTAGAGTTGTGAAGAAGAAATTTTACAAGATA---ATAATA 2171

Db 111 AATATGTGTGACCACTGTGTGTATATATTTATTAATTTTAAATTTATATATAA 170

Qy 2172 TTTTATTTATAGAAATATAGATGATTTTGTGAAAAAATTTATTGAAGTTCTTCAAGCAA 2231

Db 171 ATAATTTTAAATAAATATATATAAATTTTATATAAATAATATTTTATTAATAAATAA 230

Qy 2232 CTAACCTTTACATCAATATATATTTTGTGAAAGATTAACAATAGTTGGAATAAT 2291

Db 231 ATAAATTTTATATATATATATATATTTTAAATAATTTTAAAAATAATAATTTAAT 290

Qy 2292 TTAATGAGATCAAGAAATGAATAATAAATAAGATGCATATTTTGATTAATGGCTTATCAT 2351

Db 291 ATAATAATATATATATATATATATATATATATATATATATATATATATATATATATAT 350

Qy 2352 AATTTTCTCAGATTTTACTGGAGAGGATACAGATATATATCATCTCTCTCA-GGAGAA 2410

Db 351 TATTTATATAAAAAATTTAATATAAATTAATTTATATAATATATATATATATATATATAT 410

Qy 2411 TGCACACATTAGTTCCTTCAGAAATACCTGATATAATTTATTTTAAATATATCTCAGATTTA 2470

Db 411 TTTTAAATTTATATAATTTTAAATATATATATATATATATATATATATATATATATATAT 470

Qy 2471 TATGTTGAATTTACAAAGATGACCAAAATATATAAGAAATAGATATATATGACGAGTT 2530

Db 471 ATAATTAATAATTTATATATATATATATATATATATATATATATATATATATATATAT 530

Qy 2531 AAATGTTACAGATTTTCTTAATATATACAAAAAACTATTTGATATATGTTACTGTTTGA 2590

Db 531 ATATTTATATATATAAATTTTATATATATATATATATATATATATATATATATATATAT 590

Qy 2591 ATTTTATTAAGAATGTATCGAGCTTTTGAATCTATTTTACAAAGATTTGTTGTTTATTTGAT 2650

Db 591 ATATAATATAATTTATTAATAATTAATTTATTTATTAATAATAATTTTATATATATAT 650

Qy 2651 AGAATAAAAAACATGGTCTAAGAAATAGATTTGGTCTTAATTTGGGTTTCGCTTCCACATG 2710

Db 651 ATATATTTAAATTAATAATATATTTTAAATATATTTTATAATAATTTTAAATTTATATAT 710

Qy 2711 ATTTTGTGGCAAT-----CTTTTATCAATGAAAAACGAAAC 2747

Db 711 ATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 770

Qy 2748 AGCTTATTTTAAAGTAATCTTAATGTCAGATGAACATAATTTATACAGACAAATATAGA 2807

Db 771 AATATATAATAATATATATATTTATTTTAAATATATAATAATAATTTTATATATATATA 830

Qy 2808 AANAATGAATTTTCAATAGATTTATCTAATAATGGAATTTTAAGATATATAAAGTGAA 2867

Db 831 TTTATCTATAATTTTATTTTATTTTAAATAATAATAATAATAATAATAATAATAATA 890

Qy 2868 AAATCAACATCTTCTCTATTTGCTT---TACAGATGATTTCTTATGATCAATTTGCTAA 2923

Db 891 TATTTTATAAATTTTAAATAAATTAATAAATTAATTTTATATATATTTTAAATTTATAT 950

Qy 2924 ATCAAGAAATTTAGGTTTTTATTTCTGAGAAAGTTAAAAATAGAAATAAATCTAAAT 2983

Db 951 AATATATATATTTTACTTATTTATTAATATATATATATATATATATATTTTAAATTTAAT 1010

Qy 2984 TTAAGAANAATTTTATCTAAAAAATAAATAGTTGATTTTGTGAGAGTAATGATGTTTAA 3043

Db 1011 AATTTAAAAATCTTTTATATAAATTAATTTTATTTTCAATATATAATAATATATTTTAA 1070

Qy 3044 ATATTTTAAATATGACCCGGAAATTTTATTTTAAAGTACTTCTGGTTGATTTATTTAT 3103

Db 1071 TATATAAACTTTTAAATTTTAAATAAATTAATTTTATTTATTAATTAATAATAATATAT 1130

Qy 3104 TCCAGACAAAGATGATGTTATTTTATTAATTTTATCAATTTTATTTTATTTTCAATATA 3163

Db 1131 TATATATATTAATTTAATATATATAAATTAATTAATTAATTAATTAATTAATTAATTTAT 1190

Qy 3164 ATTTTGAACAATTAAGCTAATATTTAAAAAATGAAATTTTATGTTTTTATTTATGCTCAT 3223

Db 1191 ATTTTAAATTTTAAATTTTATATAAATTAATAAATAAATAAATAAATAAATAAATAA 1250

Qy 3224 ATATGTTTTGTTTTCAGTAGTCACAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3283

Db 1251 ATATATATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 1310

Qy 3284 TGCAGATTTTACTGCTCCATAAT---TTGATTTTTCGAATATGATTTATTAATAATTTGTA 3340

Db 1311 TATATATATTTATTTATATATATAATTAATTTATTTATTTATTTATTTATTTATTTAT 1370

Qy 3341 TTCAATTTTATAAATTTGATTTTATAAATAAATAAATAAATAAATAAATAAATAAATAA 3400


```

Db 1371 TAAATTCACATATATATATATAATAATATTTATATATTANCATTTTATATATATATATATATA 1430
Qy 3401 TTTATTAGGTATATCTCGATGTATATATATTTCAAAATGGGAAGATATGTATTTT 3456
Db 1431 TATACATTTAATAATTTTATATAAATTAATTAATATATATANACTATTTATATATATAT 1486

RESULT 13
CG750135
LOCUS CG750135 1592 bp DNA linear GSS 24-OCT-2003
DEFINITION P044-3-D09.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.
ACCESSION CG750135
VERSION CG750135.1 GI:37971278
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 1592)
REFERENCE Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
AUTHORS Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus
pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE 228835951
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@uebingen.mpg.de
Class: BAC ends.

FEATURES
source
1..1592
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/notes="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

ORIGIN
Query Match 2.1%; Score 144; DB 9; Length 1592;
Best Local Similarity 41.3%; Pred. No. 4.1e-12;
Matches 574; Conservative 0; Mismatches 810; Indels 7; Gaps 3;

Qy 2067 CCAAAGGAAAAACAATATTTGTTCTCTAGACAAAAAAGTATGGTGAAACATGTAAATG 2126
Db 127 CNNNAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 186
Qy 2127 ATCATCAAGTAGATTTGTAGAGAGATTTTACAGATAATAATATTTTATTTATATAGAAA 2186
Db 187 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNA 246
Qy 2187 ATATAGATGATTTGTTTGAAGAAATTTTGAAGTTTCTAAGCAAACTAACTTTTACATCAA 2246
Db 247 NAAAAAANANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 306
Qy 2247 ATATAAATTTTTTTGTGAAGATTAATAACAATAGTTGAAAAATTTTAATGAGGATCAAG 2306
Db 307 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 366
Qy 2307 AATATGATATATAAAGATGCATATTTGATATATGCTTATCATAATTTTTTCTCAGATT 2366
Db 367 AAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 424
Qy 2367 TTACTGAGAGGGGATACAGATATTTATCATCTCTCTCAGGAGATGCACACCATTAGTTC 2426

```

```

425 ATANNAANANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANNN 484
Qy 2427 CTTCAGAAATACCTGTATAATATTTTAAATATTTCTCAGGATTTATATGTGTAATTTCAAA 2486
Db 485 AATTAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAA 544
Qy 2487 AAGATGAGCAAAATATAAAGAAAAATAGGATATATGACGAGTTAAATGTTCAGATTAT 2546
Db 545 AAAAAAANANAAAAAANAAAAAANAAAAAATTTTTTTTTTTTTTTTTTTTATTTTATTT 604
Qy 2547 TTCCTAATATATCAGAAAAAATCTATTGATATCTACTGTTTACAGATTTTATTAAGAATGT 2606
Db 605 TATATTATTAANNAAAAATTTTATTTATATATTTTATTTATTAATTAATAATTT 664
Qy 2607 ATCGAGCTTTTGAATACTATTTCAAAAGATTTGTGTTTATTTGATAGATAAAAAACATGG 2666
Db 665 ATAATTTATTTATTAATATTTTAAATAATTTTTTTTTTTTATTTATATATATATTTTATTTA 724
Qy 2667 TCTAAGAAATAGATTTGGTTCATTTGGGTTTCGCTTCCACATGATTTTGTGGCAATTC 2726
Db 725 TTAATAATATTTTATATAAATAATTAATAATAATTTATATATANTAAAAATAATTTATTAAT 784
Qy 2727 TTTATCAAAATGAAACGAAACAGCTTATTATTTAAAGTAATCTAAATGTCCAGATGAAT 2786
Db 785 TTTTATNAATTTATATATAATAATATATATTTTAAATTAANTTTTANATAATAATTTATTAAT 844
Qy 2787 ATTATACAGACAAATTTAGAAAAATATGAATTTTCAAAATAGATTTCTAAATATGAAA 2846
Db 845 TTTTATTAATAAATAATTTTATATAATAAATTTTAAATTTTAAATTTTATTAATTTTAT 904
Qy 2847 TTTAAGATATATAAGTGGAAAAAATCAACATCTTCTCTATTTGCTTTACAGATGATTC 2906
Db 905 ANATTATAATTTTANAAAAATATAATAATAATAATAATAAATAATTTTATTTATTTAATA 964
Qy 2907 TATTGATGAATTCGCTAAATGCAAGAAATTTAGGTGTTTTTATTGCTAGAAAGTTAAAAAT 2966
Db 965 TATATTNNATTT--TTTATTTATTTATTTTATTTTAAATTTATATATTTATTTATATAT 1022
Qy 2967 AGAAAAATAATCTAAATTTTAAAGAAATTTACTAAAAAATAAATAAGTTGATTTGTGA 3026
Db 1023 TTATATTTTATTTTAAATATATNATTTAAATATATAAATAATTAATAATTTATTTATTTT 1082
Qy 3027 GAGTATGATGTGTTAAATTTTAAATATGACCCGGAATTTTATTTTATTTTAAAGTACTTC 3086
Db 1083 AATTAATTTATATTTTATTTTAAATTTATATATATTTTATTTTATTTATTTATTTAATA 1142
Qy 3087 TGGTTGATTTATTTTATTTCCAGAGCAAAAGTATGTTATTTTATTTTATGAAATTTA 3146
Db 1143 TTTATTTTAAATTTTATTTTATTTTATTTATTTTATTTTAAATTTTATTTATATATATTT 1202
Qy 3147 ATTTTATTTTATATAAAAAATTTTGAAGAACTAAGCTAATAATAAAAAAATGAAATTTTATG 3206
Db 1203 TTTATATTTTAAATTTTAAATTTAAATTAATTAATAATAATAATAATAAATAAATAAATA 1262
Qy 3207 TTTTATATATGCTATATATTTGTTTGTTCAGTAGTCAAGTATGTTGTTGTTGTTGAATA 3266
Db 1263 TATTTAATATTTATATAATTTAAATTTTATTTTATATATTTTATTTATTTTATTTATATAN 1322
Qy 3267 AATTTTGAAGATTTATTTGAGATTTTACTGCTCCCATTAATTTGGGATTTATGCAATAATG 3326
Db 1323 TATTAATAATATAAATTTAAATTAATAATATATATTTATTTATTTTAAATTTATATAATATA 1382
Qy 3327 TATTAATAATTTGATTTTATTTAATAATTTGATTTAATAAAAAATTAATAAATAGTATCTTT 3386
Db 1383 TTTTATTTTATTTATATAAATTTATTAATAAATTTAAATTTTATTAATTTTATTTTAAATTT 1442
Qy 3387 TTTAGTTTTTTTAGTTTTTATAGCTATATCTGCATTTGCTATATTTATTTTCAAAATGGGAAGAT 3446
Db 1443 TTTTATTTTATATATTTATTTTATATATATATATATATATATATATATATATATATATAT 1499
Qy 3447 ATTTGATTTTTT 3457
Db 1500 ATTTAATTTAT 1510

```


| | | | | | |
|------------|--------------------------------------------------------------------|---------------------------|-------|--------------------|-----------------|
| RESULT 15 | CL068807 | 1242 bp | DNA | linear | GSS 31-DEC-2003 |
| LOCUS | CH216-115B3 | Sp6.1 | CH216 | Xenopus tropicalis | genomic clone |
| DEFINITION | CH216-115B3 | _genomic survey sequence. | | | |
| ACCESSION | CL068807 | | | | |
| VERSION | CL068807.1 | GI:40524720 | | | |
| KEYWORDS | GSS. | | | | |
| SOURCE | Xenopus tropicalis | (western clawed frog) | | | |
| ORGANISM | Xenopus tropicalis | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| | Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; | | | | |
| | Xenopodinae; Xenopus; Silurana. | | | | |
| REFERENCE | 1 (bases 1 to 1242) | | | | |
| AUTHORS | Krematzi, C., Carter, J., McPherson, J., Warren, W., Graves, T., | | | | |
| | Mardis, E. and Wilson, R. | | | | |
| TITLE | A physical map of the xenopus tropicalis genome | | | | |
| JOURNAL | Unpublished (2003) | | | | |
| COMMENT | Contact: Richard K Wilson | | | | |
| | Genome Sequencing Center | | | | |
| | Washington University School of Medicine | | | | |
| | Email: submission@watson.wustl.edu | | | | |
| | Insert Length: 175000 Std Error: 0.00 | | | | |
| | Seq primer: Sp6 ATTAGTGACACTATAG | | | | |
| | Class: BAC ends | | | | |
| | High quality sequence start: 7 | | | | |
| | High quality sequence stop: 57. | | | | |
| FEATURES | Location/Qualifiers | | | | |
| source | 1. 1242 | | | | |
| | /organism="Xenopus tropicalis" | | | | |
| | /mol_type="genomic DNA" | | | | |
| | /strain="Nigerian frog" | | | | |
| | /db_xref="taxon:8364" | | | | |
| | /clone="CH216-115B3" | | | | |
| | /sex="male" | | | | |
| | /cell_line="Stock 248 F7A2, inbred N7" | | | | |
| | /clone_lib="CH216" | | | | |
| | /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis | | | | |
| | BAC library" | | | | |
| ORIGIN | | | | | |
| | Query Match 2.0%; Score 143.2; DB 9; Length 1242; | | | | |
| | Best Local Similarity 46.9%; Pred. No. 5.5e-12; | | | | |
| | Matches 545; Conservative 0; Mismatches 613; Indels 4; Gaps 3; | | | | |
| Qy | 2484 CAAAGATGACGCAAAATATTAAGAAATATAGGATATATGACGGTTAAATGTTACAGAT | 2543 | | | |
| Db | 79 CACAAAATAAAGAAAGAAAAAAGAAAAAATTTTTTTATAAAATTTTTTTAT | 138 | | | |
| Qy | 2544 TATTTCCTAATATATACGAAAAAACTATGTAATGTACTGTTAGAAATTTTATTAGAA | 2603 | | | |
| Db | 139 TTTTATTATTTTATATATTTTATATATATTTTAAATTTTTTTATTTATTTATTTATTT | 198 | | | |
| Qy | 2604 TGTATCGAGCTTTTGAATACATTTTACAAAGATGTTGTTTATTTGATAGATAAAAAACA | 2663 | | | |
| Db | 199 AATATTTATATATTTAAATTTATTTTATATATTTATTTATATAATTTTA-ATTTAAATTTA | 257 | | | |
| Qy | 2664 TGGTCTAAGAAATAAGATTGCGTCTTAATTCGGTTTCGCTTCCACATGATTTTGTGGCAAT | 2723 | | | |
| Db | 258 TTTTATATTTTTCGTTATTTATTTATATTTTAAATTTTATTTATTTATTTATTTATTTT | 317 | | | |
| Qy | 2724 TCTTTTATCAAAATGAAAAACGACGCTTATTTATTTTAAGTAAATCTAAATGTCCAGATGA | 2783 | | | |
| Db | 318 TATTAAATTTTATATATATTTATTAATTTATTTATTTATTTATTTTATTTTATTTATTT | 377 | | | |
| Qy | 2784 ACTATTTTATACAGCAATATAGAAAAATATGATTTTTCAAAATAGATTAATCTAAATATGG | 2843 | | | |
| Db | 378 AATATAATTTTTTTTAAATATTTAAATTAATTTATTTTATTAATTTATTTTAAAAAAA | 437 | | | |
| Qy | 2844 AAATTTTAAGATATATAAAGTGGAAAAATCAACATCTTCTCCTATTCTCTTTACAGATGA | 2903 | | | |


```

RESULT 18
CL110653          1594 bp   DNA       linear       GSS 05-JAN-2004
LOCUS             ISB1-53P23 Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-53P23,
DEFINITION        genomic survey sequence.
ACCESSION         CL110653
VERSION           CL110653.1 GI:40604288
KEYWORDS          GSS.
SOURCE            xenopus tropicalis (western clawed frog)
ORGANISM          Xenopus tropicalis
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
                  Xenopodinae; Xenopus; Silurana.
REFERENCE         1 (bases 1 to 1594)
AUTHORS           Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
                  Mardis, E. and Wilson, R.
TITLE             A physical map of the xenopus tropicalis genome
JOURNAL           Unpublished (2003)
COMMENT           Contact: Richard K Wilson
                  Genome Sequencing Center
                  Washington University School of Medicine
                  Email: submissions@watson.wustl.edu
                  Insert Length: 75000 Std Error: 0.00
                  Seq primer: Sp6 ATTAGGTGACACTATAG
                  Class: BAC ends
                  High quality sequence start: 390
                  High quality sequence stop: 470.
FEATURES          Location/Qualifiers
                    1..1594
                        /organism="Xenopus tropicalis"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:8364"
                        /clone="ISB1-53P23"
                        /clone_lib="ISB1"
                        /note="Vector: phelobAC11; ISB-1 Xenopus tropicalis BAC
                        Library Segment 1"
ORIGIN
    Query Match      2.0%; Score 142.6; DB 9; Length 1594;
    Best Local Similarity 45.2%; Pred. No. 6.7e-12;
    Matches 608; Conservative 0; Mismatches 729; Indels 8; Gaps 2;

Qy 2961 AAAAAAGAAAATAAATCTAAATTTAAAGAAATTTACTAAAGAAATTTAAATAAGTCTGATT 3020
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 229 AAAAAAGAAAAGAGAGAAAACCAAGAGAAAAGAGCAAAAAGAAAATTTTTTTTTT 288
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3021 TTGTGAGAGTAAGTATGTTTAAATTTATTTAAATATGACCCGGAATATTTTATTTTAAAG 3080
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 289 TTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTATTTATTTTTTTTTTTTTT 348
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3081 TACTCTGTTGATATTTTATTTCCAGAGCAAAAGATGATTTTATTTATTTATTTTATG 3140
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 349 TTTTATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 408
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3141 AATTAAATTTTATTTTATATAAAATTTTGAAGAACTAAGCTAATATTAATAAAGAAAT 3200
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 409 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 468
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3201 TTATTTGTTTTTATGTCATATTTATTTGTTTTCAGTAGTCACAAAGATGTTTGT 3260
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 469 TTTATTTTATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 528
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3261 GAAATAAATTTGAAAGATATTTGCGAGATTTTACTGCTCCCAATATTTGGATTTATGCA 3320
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 529 TTTTATATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 585
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3321 ATATGATTTATATTTCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3380
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 586 ----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 641
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3381 ATCTTTTTTAGTTTTTTAGTTTTTATTTAGTATATCTGCAATGATATTTATTTCAAAATGGG 3440
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 642 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 701
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3441 AAGATATTTGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3499
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 702 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 761
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3500 CAAAACAAGGTGGTGGCTTTTATGAACATCTCCAGTAAATACCACCTACCAATATATAGT 3559
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 762 TTAATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 821
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3560 TTCAATTCGGTAAATCTTTGCACCTATATAAATAAATAAATAAATAAATAAATAAATAAATA 3619
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 822 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 881
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3620 TCTTGTCTTTATATCCGATCTATTTTAAAGTGGATCGAGAAATTCGTTAGTTTATCGCTAGCAAT 3679
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 882 TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 941
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3680 ATTAATTTATATGCTTTGTTATGGAGATATATAGGTGGAAATTTGCTGGATATAAAGCT 3739
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 942 TTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1001
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3740 AATAGTAATATTTGTAATACTACTTATTTTAAATCTCAATGCTTTTACCAGTAAT 3799
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1002 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1061
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3800 TTTGGCTGTTTATTAATTTCTAGAGAAATCAAGTAACGAAGCTAGATTTTATTTATTTATCAAGG 3859
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1062 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1121
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3860 AAGTATTTGATAAAGTATTTAGAAAAAATAATTTTATTTGGATGGAATATCCGAATATTC 3919
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1122 TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1181
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3920 AGTTACGGAACTTGGCTCGGAAGTCAATCAGGCTATATATCATTTTATTAATCAAGG 3979
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1182 AATTTTATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTATTT 1241
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3980 AATAGTTGGGTGATTTTACTGATGTTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTG 4039
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1242 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1301
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4040 AGTTAAATGGGAAACAGCAGCTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4099
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1302 TTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1361
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4100 AACAAATAGATCCGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTG 4159
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1362 ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1421
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4160 TAATATAAATTTTAAAGGATATGGAGACAAAATAAGTAAATTTTATTTTATTTTATTTTATTTG 4219
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1422 GTTTTATTTATTTGATATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1481
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4220 TACCAATTTATTAATGTCAGATTTATCTTGATAAATGTTATTAACAGATTTATTTATTAACCAA 4279
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1482 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1541
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4280 CATATACTAATTTAGAGGTTATTTCT 4304
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1542 TTTTATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1566
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 19
CL118721/c
LOCUS             ISB1-72J8 T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-72J8,
DEFINITION        genomic survey sequence.
ACCESSION         CL118721
VERSION           CL118721.1 GI:40612356
KEYWORDS          GSS.
SOURCE            Xenopus tropicalis (western clawed frog)

```


[illegible]


```
QY 2404 AGGAGATGCACACCATTAGTCTCTTCAGATACCTGTATATAATTAATTTAAATATCTCA 2463
Db 1193 TAATTTTATATACTATTTATATATTTTAAATATATATATATATATATATATATTTT 1134
QY 2464 GGATTTATATGTTGAAATTTTCAAAAGATGACGAAAAATATAAGAAAAATAGGATATATGA 2523
Db 1133 ATATTTAAATATATTTTATTTTACTTATATATATATACATATATCTTTATATATTTTCAT 1074
QY 2524 ACGAGTTTAAATGTTTACAGATTTTCTCTAATATATATACAGAAAAAATCTATTGATAATGACT 2583
Db 1073 TTATATATATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1014
QY 2584 GTTTAGAAATTTTATAGAAATGATCGAGCTTTTGAATACATATTTACAAAGATTTGTTT 2643
Db 1013 TTATATTTTATATATTTTATTTTATTTTAAACCTTTTATTTTATATATATA-----TATTTT 959
QY 2644 TATTGATAGAAATAAAAACATGGTCAAGAAATAGATTTTGGTCTCAATTCGGTTTCGCTT 2703
Db 958 ATTAAAAAATAATTTTTTTTTTATATTTTAAATATATATTTTCTTTTATTTATATATATTT 899
QY 2704 CCACATGATTTTGGGCAATTTCTTTATCAAAATGAAACGAAACAGCTTTATTTTAAAG 2763
Db 898 TTTTATTTTATTTTAAATATATCTATATTTTAAATTTATTTTATTTTATTTTATTTTATTT 839
QY 2764 TAATCTAAATGTCAGATGAATTTTATATACAGCAATTTATAGAAAAATATGAATTTTCA 2823
Db 838 TTTTATATANATTTTATATATTTTAAATTTTATTTTATATATTTTATTTTCTTTTATATAA 779
QY 2824 AATAGATTTATCTAAATATGAAATTTTAAAGATATATAAGTGGAAAAAATCAACATCTTCT 2883
Db 778 TAAATTTTATTTTATATA-----ATTTTATATATATATATTTTATATTTTATTTTAAATA 722
QY 2884 CCTATGCTTTACAGATGATTTCTTATTTGATGAATGCTAAATGCAAGAAATTTAGGTTTT 2943
Db 721 AATTTTATATTTTATTTTATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 662
QY 2944 TTATTCTAGAAAGTTTAAAAATAGAAAAATGAAATTAATCTAAATTTTAAAGAAATTTACTAAA 3003
Db 661 TATTTTAAATATTTTATTTTATATATATATATTTTATTTTAAATTTTATTTTATTTTATTTT 602
QY 3004 AATAAAAATAGTTGATTTTGTGAGAGTAAATGATGTTTAAATTTTAAATATTTTAAATATGACCGG 3063
Db 601 AATATATATTTTATATATTTTATATATTTTATTTTATTTTATATATTTTATATATATTTT 542
QY 3064 AATATTTTATTTTAAAGTACTCTGTTGATTTATTTTATTTCCAGAGCAAAAGTATGTAT 3123
Db 541 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 486
QY 3124 TTTTATTTTATTTTATGAAATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3183
Db 485 TAAAAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 426
QY 3184 TATTAAAAATGAAATTTTATTTTATTTTATTTGTCATATATTTGTTTGTTCAGTAG 3243
Db 425 AAAAAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 366
QY 3244 TCACAGATGTTTGTGTTGAAATTTTGAAGATTTTTCGAGATTTTTCGCTGCCCA 3303
Db 365 TAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 306
QY 3304 TAATTGGATTTTGGCAATGATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3363
Db 305 TTTTATATATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 246
QY 3364 AAAAAATTAATAATA 3378
Db 245 AAAAAAATTAATAA 231
```

RESULT 24
AG333983/c
LOCUS

DEFINITION Mus musculus molossinus DNA, clone:MSMg01-124L03.T7, genomic survey sequence.

ACCESSION AG333983

VERSION AG333983.1 GI:47907293

KEYWORDS GSS.

SOURCE Mus musculus molossinus

ORGANISM Mus musculus molossinus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

BAC end Sequences of Library MSMg01

Unpublished

2 (bases 1 to 1943)

Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

Direct Submission

Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan

1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,

Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the mouse BAC library MSMg01. For BAC

library availability, please contact Kuniya Abe (abe@rtc.riken.jp).

Tsukuba Institute Bio Resource Center, Research (RIKEN) 3-1-1

The Institute of Physical and Chemical Research (RIKEN) 3-1-1

Koyadai, Tsukuba, 305-0074 Japan

phone: 81-298-36-9189, fax: 81-298-36-9199

e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : T7

LIBRARY : pBACe3.6

Vector : EcoRI

R.Site 1 : EcoRI

R.Site 2 : EcoRI

Location/Qualifiers

1. 1943

/organism="Mus musculus molossinus"

/mol_type="genomic DNA"

/sub_species="molossinus"

/db_xref="taxon:57486"

/clone="MSMg01-124L03.T7"

/sex="male"

/tissue type="mixture of kidney and spleen"

/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 2.0%; Score 137.8; DB 9; Length 1943;

Best Local Similarity 45.7%; Pred. No. 3.6e-11;

Matches 828; Conservative 0; Mismatches 947; Indels 35; Gaps 9;

QY 2095 TAGACAAAAAAGTATGGTGAACATGTAATGATCATCAAGTAGAGTTGTAAGAAGAT 2154

Db 1911 TATTAAGATATAATATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1852

QY 2155 -TTTACAGATAATATATTTTATTTATAGAAAAATAGATGATTTGTTTGAAGAAATTA 2213

Db 1851 AATAATTAATATAATAAATGTAATGTAATGATGATGATGATGATGATGATGATGATGATG 1792

QY 2214 TTGAAGTTCTTAAGCAAACTAACTTTTACATCAAAATAATTAATTTTGTGAAAGATTAA 2273

Db 1791 TATAATAAATTTTGTATTTTATATCATATTCGCAAAATACATAGATTTTATTTATTAAT 1732

QY 2274 AACAAATAGTTGAAAAATTTAATGA--GGATCAAGAAAAATGAATAATAAAAAAGATGCAT 2331

Db 1731 AAATGATTATTAATTAATTTTAAATATATGAATAGATATAAAATGCAATAAATAATTT 1672

QY 2332 ATTTGATAATGGCTTATCATATTTTCTCAGATTTTCTCGAGAGGGATACAGATATTA 2391

Db 1671 ATATTATAGTAACGTTTCATGAATTAATAAATTCGTAATTTATTTATATAGTATATAGATA 1612

QY 2392 TCATCTTCTCAGGAGGAATGCACACCATTTAGTTCTCTCAGATACTCTGATATATTTT 2451

Db 1611 TAAT-TTATATACTAATTAATAATTTAATGTTTAAATTTATATATTTTAAATATAGAGATA 1553


```
Qy 2452 TAAATATCTCAGGATTTATATGTTGAATTTTCAAAAGATGAGCAAAATATAAAGAAAA 2511
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1552 TATATATATATAGTAGTTATATATATATATATATATATATATATATATATATATATAT 1493
Qy 2512 TAGGATATATGAACGAGTTAAATGTTACAGATTTATCTCTAATATATCAGAAAAAATCTAT 2571
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1492 AGAACATTTATATATATATATATATATATATATATATATATATATATATATATATATAT 1433
Qy 2572 TGATAATGTACTGTTTGAATTTTATTAAGAAATGTATCGAGCTTTTGAATCTACTATTTACA 2631
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1432 TATAAAATGAATAAATTAATAAATACTACTTTATCTTAATAATTAATAATAATTTAT 1373
Qy 2632 AAGATTGTTTATGTATAGATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2691
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1372 TATATATTTGTGATAATGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1314
Qy 2692 TGGGTTTCGCTCCACATGATTTTGTGCAATTTCTTTTATCAATGAAAAAGAACAGCT 2751
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1313 ATGATATGAAATTTATATAATTTTTGTAAATTAATAATAATAATAATAATAATAATAATA 1254
Qy 2752 TATTTATTAAGTAAATCTAAATGTCCAGATGA-----ACTATTTATACAGACAAT 2801
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1253 AAGATGTATTAATAATATATATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1194
Qy 2802 TATGAAAAATATGAATTTTCAATPAGATTTCTAAATATGGAATTTTAAAGATATATAAA 2861
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1193 AATGAATAAATAAATAATCAAAAAATAATTCATAGTATAATTTATTAATAATAATAATAA 1134
Qy 2862 GTGGAATAAATCAACATCTTCTCTATTTGTCTTTACAGATGATTCATTGATGAATGCT 2921
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1133 TAATAATATATAATTTTATAATATATAATAAATAATAATAATAATAATAATAATAATAATCA 1074
Qy 2922 AAATGCAAGAAATTTAGTGTTTTTATTTGTAGAAAGTTTAAATAATAGAAAAATAAATCTAA 2981
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1073 GTAATTTATATTAATGTATAAATAATTAATGTAATAATTTAATTTGAATTAATAATA 1014
Qy 2982 ATTTAAAGAAATTTACTTAAATAAATAAATAGTTGATTTTGTGAGAGTAATGATGTTT 3041
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1013 ATATTAATAATATATAAATAATTAATAATTTATATAATAATAATAATAATAATAATAATAAT 954
Qy 3042 AAATATTTAAAT---ATGACCGGAATATTTTATTTTAAAGTACTTCTGCTGGATTAAT 3098
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
953 TATATATTAATGAAATAATAATATCATCTTTAAATAATAATAATAATAATAATAATAATAAT 894
Qy 3099 TTTATCCAGAGCAAAAGTATGTTTTTATTTATTTTATGAAATTTTATTTTATTTTATCAT 3158
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
893 TAGATAATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 840
Qy 3159 ATAAAAATTTTGAAAACTAAGCTAATAATTAATAATAATAATAATAATAATAATAATAATA 3218
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
839 ATATATATATAGATAGATATAGTAATTTGTAATAATAATAATAATAATAATAATAATAATA 780
Qy 3219 TCTATATATGTTTGTGTTTCAAGTATGCAAGTATGTTTGTGTAATAATAATAATAATAATA 3278
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
779 TATATATAATAATAAGTATTTTATATATAATAATAATAATAATAATAATAATAATAATAATA 720
Qy 3279 TTTATTCAGATTTTACTGCTCCCATATTTGGATTTATGCAATAATGTAATAATAATAATA 3338
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
719 ATGTTGAGAAATATATAATTTTAAATAATAATAATAATAATAATAATAATAATAATAATAACT 660
Qy 3339 TATTCATTTATAAATATGATTTATAAAAAATTAATAATAATAATAATAATAATAATAATAATA 3398
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
659 AATTTATATAGATATTTGTTAAACATAATAATAATAATAATAATAATAATAATAATAATAATA 600
Qy 3399 GTTTTATTAGGTATATCTGCAATGTTGATATTTTCAAAATGGAAGAATTTGATTTTATA 3458
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
599 ATATAAATACA-----ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 549
Qy 3459 GACAGACCTTATAGGACTAGACTATCTTATAACAGCGCTCAAAACAGGTTGGTTGGC 3518
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
548 TATATTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 489
```

```
Qy 3519 TTTATGAACATCTATCCTACGTTAAATACCACTACAATTAAGTTTCAATTCCTGTTAATCTTT 3578
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
488 TTAATTTATTTTAAATATTTATATATTTAAATATTTATGAGTATAAATATAAATTTATTTATA 429
Qy 3579 GCACCTTATAAAAAATAAATGCAACAATTTTTTTTCTGCTGCTTTTATATACCGATC 3638
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
428 AATTTATTAATTTATATATAGGAAAAAAGAATTTAAATTAATTTATTTGATATATATATAT 369
Qy 3639 TATTTAAGTGTGATCGAGAAATTTGTTAGTTTATCGCTAGCAATATTAATTTATATGCTTTGTTA 3698
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
368 TTAATTTATTAATTTAGATTTAAATTAATATAATAATAATAATAATAATAATAATAATAATC 309
Qy 3699 TGGAGATATATAGTGTGGAATTTTGTCTGGTAAATAAAGCTAAATAGTAATTTTGTGTAATA 3758
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
308 AATAAATAAGTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 249
Qy 3759 CTACTTATTTTAAATACTGA--ATTGCTTTTACCAGTAATTTTGGCTGCTTTTATTAATTT 3816
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 ATTTATTAATAATAATAATAATTTATTAATTTTAAATAATATAATAATAAATAAATAAATAA 189
Qy 3817 CTAGAGATCAAGTAACGAGCTAGATTTTATTTTATCAAGGAAGTATTGATAAAGTAT 3876
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
188 CTTTAAAGAAACACACACACACACACACACACACACACACACACACACACACACACACACAC 129
Qy 3877 TAGAAAAACA 3886
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
128 AAAAAAANA 119

RESULT 25
LOCUS CL082000
DEFINITION CH216-165P18 RM4.1 CH216 Xenopus tropicalis genomic clone
CH216-165P18, genomic survey sequence.
ACCESSION CL082000
VERSION CL082000.1 GI:40537913
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1981)
AUTHORS Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM4 ctcaagggtcgcgtcgagc
Class: BAC ends
High quality sequence start: 265
High quality sequence stop: 497.
FEATURES
source 1. 1981
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-165P18"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/notes="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"

ORIGIN
Query Match 2.0%; Score 137.6; DB 9; Length 1981;
Best Local Similarity 43.4%; Pred. No. 3.8e-11;
Matches 674; Conservative 0; Mismatches 876; Indels 4; Gaps 2;
```



```
QY 2751 TTATTTATTTAAGTAATCTAAATGTCAGATGAAGTATTTATATACAGACAATATATAGAAA 2810
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
26 TTTTTTCTTTTTTATCTTACTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTAAATAAATTTTTNT 85

QY 2811 ATATGAATTTTCAAATAGATTAATCAATAATGGAATTTAAAGATATATATAAGTGGAAAAA 2870
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86 TTTTTTTTTTTTTATATATAT - TTTATTTTTTTTTTATTTTTTGTATTTATTTATTAATTAATTA 144

QY 2871 ATCAACATCTCTCCTATGCTTTTACAGATGATTCATTTGATGAATTCCTAAATGCAAG 2930
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 ANNNCCCTTTTTTTTTTTTTTTGGGGGANNNNNNNNNNNANNTAANTTAATATTTTA 204

QY 2931 AAATTTAGGTTTTTATTTCTAGAAAGTTAAAAATAGAAAAATAATCTAAATTTAAAGA 2990
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205 TTAAGACCTTTTTTTTTTTTTTTTGAGAGAGANNNAATAAATTAATAAATATTTTTT 264

QY 2991 AATTAATTAACAAAAATAAAATAGTTGATTTGTGAGAGTAATGTATGTTTAAATTAATTT 3050
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTATATATTTTTTTTTTTTTT 324

QY 3051 AAATATGACCCGGAATTTTATTTTAAAGTACTCTCTGGTTGATTTATTTATTTCCAGAG 3110
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
325 TTTTTTTTTTTTTTTATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 384

QY 3111 CAAAAGTATGATTTTATTAATTTTATGAATTAATTTTATTTATCATATATAAATTTTGG 3170
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
385 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTT 444

QY 3171 AAAAATAAGCTAATATTAATAAAATGAAATTTTATTTGTTTTATTTATGCTATATTAATGT 3230
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
445 TTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 504

QY 3231 TTTGTTTCAGTAGTCAAGATATGTTTGTGAATAAATTTGAAAGATTAATTTGCAGAT 3290
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
505 TTTTAATTTTTTTTTTTTAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 564

QY 3291 TTTACTGCTCCCAATAATTTGGAATATGCAATAATGATTAATTAATTTGATTCATTATTA 3350
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
565 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTATGTTTTTTT 624

QY 3351 AATATGATTAATAAAAAATAAAAATAGTATCTTTTTTTAGTTTTTTTTTAGTTTTATTAGT 3410
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
625 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAGTTTTATTAATTTTTTTAT 684

QY 3411 ATATCTGCATGTATATATCAAAAATGGGAAAGATTTGTAATTTTTTAGACAGACACCTT 3470
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
685 TTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTATTTATTTTTTTTTTTTTTTTTTTTTTGTTTTTT 744

QY 3471 ATAGGACTAGACTATCTTATACAGGGCTCAAAACAGGTGGTGGCTTTATGAACATAT 3530
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
745 TATTTTTTTTTTATTTTTTTTTTATTTTTTTTTTTAAATTTTTTATTTTTTTATTTGTTTTATTTT 804

QY 3531 CCTACGTTAAATACCACATAATATAGTTTCAATCCGTTAATCTTTGCACTTATAAAA 3590
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
805 GTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATA 864

QY 3591 AATAAATGCAACAAATTTTTTTCTGTGTCTTTTATACCGATCTAATTTAAGTGGGA 3650
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
865 TTCTTTTTTTTTTATTTGTTATTTATTTTTTTATTTTTTTTTTTAATATTTTTTTTTTTTT 924

QY 3651 TCGAGAAATGCGTAGTATTCGCTAGCAATTAATTAATATATGCTTGTATGGAGATATATA 3710
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
925 TTTTTTTTTTATTTTTTTTTTTTGTATTTTTTTTTTTTTTTATTAATTAATTTATTTATATAT 984

QY 3711 GGTGGAAAAATTTGCTTGGATAAAAAAGCTAATAGTAATATTTGTAATACTACTATTTATTT 3770
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
985 TTTTTTTTTTATTTTTTATTTGTAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1044

QY 3771 TTAATACTGAAATGCTTTTACCATGAATAATTTGCTGTTTATTAATTTCTAGAGAAATCAAGT 3830
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1045 TTTTTTTTTTTTTTTTATTAATTAATTTTTTTTTTTTTTTTTTTTTTTTTTTTATAATTTTTTTTTT 1104
```

```
QY 3831 AACGAGCTAGATTTATTTATTTATCAAGGAAGTATTTGATAAGTATTTAGAAAAAATATTT 3890
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1105 TTTTTTTTATTTATTTTTTTTTTTTTTATATTTTTTTTTTATTTTTTTATTTATTTATTTATTTG 1164

QY 3891 TTATTTGGATATGGAATATCCGAATATTCAGTTTACGGGAACTTGGCTCGGAAGTCATTCA 3950
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1165 TTATTTATTTTTTTTTTTTATTTTAAATTTATTTATTTTATTTTATTTTATTTATTTTACTTT 1224

QY 3951 GGCTATATATCAATTTTTTTTATAAAATCAGGAATAGTTGGGTTGATTTTACTGATGTTTTCT 4010
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1225 GTATTTTATTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTATTTTTTTATTTGTTTTTTTTTTT 1284

QY 4011 TTTTTTATGTTATAAAAAAAGTTATGGAGTTAATGGGGAACACAGCACTATTTTTATTTT 4070
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1285 TATTTTTCCTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTATTTTATTTTCCTTTATTTT 1344

QY 4071 ACATCAATTAGCCATATTTTTTCATATATGAACAATAGATCCGATTTATTTATTTATTTAGTA 4130
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1345 TATTTTTTTTTTATTTTCTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1404

QY 4131 CTATTCCTTTCTCAATAGGTATTTTGGAAATAATAAAATTTTAAAGGATATGGAGACA 4190
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1405 TTATTTTGTATTAAATTAATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1461

QY 4191 AAAAATGAATGATTTAAATTTTCACTTATGTAACCAATTTATTAATGTCCTCAAGATTTCTTGA 4250
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1462 TTTTATGTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTATTTGATTTGTTTTTTTTTTTTTTAT 1521

QY 4251 TAAATGTTATTAACAGTATTTATTAACCAACATATACTAATTTTAGAGGTTTATTTCT 4304
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1522 TTATTTTTTTTTTTTTTTTATTTTTTTTTTTTATATATATTTTTTTTTTTTATTTTTTTAT 1575
```

Search completed: April 29, 2005, 15:55:26
Job time : 19346 secs